# The Two-tier Stochastic Frontier (2TSF) 

framework: Theory and Applications, Models and Tools.

## Alecos Papadopoulos

PhD Thesis.

November 2018.

vol.1: MAIN TEXT

EXONH

The Two-tier Stochastic Frontier (2TSF) framework: Theory and Applications, Models and Tools.

Alecos Papadopoulos

# The Two-tier Stochastic Frontier (2TSF) framework: Theory and Applications, Models and Tools. 

PhD Thesis, vol.1: Main Text.
Alexandros (Alecos) Papadopoulos
papadopalex@aueb.gr

## Athens University of Economics and Business (AUEB), Greece.

## Supervising Committee:

Plutarchos Sakellaris (supervisor), Professor, AUEB
Efthymios (Mike) G. Tsionas, Professor, AUEB
Elias Tzavalis, Professor, AUEB.
--

## PhD defense Scientific Committee:

Plutarchos Sakellaris
Efthymios (Mike) G. Tsionas
Elias Tzavalis
Tryphon Kollintzas, Professor, AUEB
Stylianos (Stelios) Arvanitis, Assistant Professor, AUEB
Subal C. Kumbhakar, Distinguished Professor of Economics, SUNY at Bighamton, USA
Christopher F. Parmeter, Associate Professor, University of Miami, USA.

This PhD thesis was presented to, examined and accepted by the Scientific Committee, on November 2, 2018. The Philosophical Diploma in Economic Sciences was awarded on November 7, 2018.

## Contents

ПЕРІАНЧН ГТА ЕИАНNIKА ..... i-xxxi
INTRODUCTION \& OVERVIEW ..... 1
Part A. REVIEW \& ANALYSIS.
Chapter 1: Literature ..... 7
Chapter 2: Structural foundations ..... 19
Part B. ECONOMETRICS \& STATISTICS.
Chapter 3: Independence and Exogeneity ..... 39
I. The Exponential 2TSF specification ..... 40
II. The Half-normal 2TSF specification ..... 45
III. A Corrected OLS/Method-of-Moments estimation method for non-closed-form 2TSF specifications. ..... 65
IV. The semi-Gamma 2TSF specification ..... 79
V. The Generalized Exponential 2TSF specification. ..... 101
Chapter 4: Dependence and Endogeneity ..... 115
I. The Correlated Exponential 2TSF specification. ..... 115
II. Accounting for regressor endogeneity using Copulas ..... 134
Part C. ECONOMIC MODELS \& APPLICATIONS.
Chapter 5: A Targets-based Nash Bargaining 2TSF framework, with a wage-determination model under productivity uncertainty ..... 157
Chapter 6: The contribution of management to production, the wrong-skewness problem, and a 2TSF model to measure them. ..... 195
Part D. SUPPORTING MATERIAL
References ..... 251
Technical Appendix (separate volume) ..... 261-528

#  

А $\lambda \varepsilon ́ \xi \alpha v \delta \varrho о \varsigma(А \lambda \varepsilon ́ \kappa о \varsigma) ~ П \alpha \pi \alpha \delta o ́ \pi о v \lambda о \varsigma$

## $\Delta \mathrm{I} \triangle \mathrm{AKTOPIKH} \Delta \mathrm{IATPIBH}$

# To $\Delta i ́ \pi \lambda \varepsilon v \varrho o ~ \Upsilon \pi o ́ \delta \varepsilon ı \gamma \mu \alpha \Sigma \tau о \chi \alpha \sigma \tau ı \kappa o v ́ \Sigma v v o ́ g o v: ~$  

(The Two-tier Stochastic Frontier (2TSF) framework: Theory and Applications, Models and Tools)

## Nó́ $\mu$ @̣tos 2018


 $\sigma v v \varepsilon \delta \varrho i ́ \alpha \sigma \eta ~ \tau о v ~ \tau \mu \eta ́ \mu \alpha \tau о \varsigma ~ O \iota к о v о \mu \iota \kappa \omega ́ v ~ E \tau \iota \sigma \tau \eta \mu \omega ́ v ~ \tau \eta v ~ T \varepsilon \tau \alpha ́ \varrho \tau \eta ~ 7 ~ N o \varepsilon \mu \beta \varrho i ́ o v ~ 2018 . ~ H ~$


## Пع@íג $\eta \psi \eta$


 $\varepsilon \nu \tau \alpha ́ \sigma \sigma \varepsilon \tau \alpha \iota \sigma \tau 0 \pi \varepsilon \delta$ ío $\tau \eta \varsigma \Sigma \tau 0 \chi \alpha \sigma \tau \iota \kappa \eta \jmath_{\varsigma} A v \alpha ́ \lambda v \sigma \eta \varsigma \Sigma v v o ́ \varrho o v$ (Stochastic Frontier Analysis /












 $\varepsilon \xi \alpha \varrho \tau \eta \mu \varepsilon ́ v \eta \mu \varepsilon \tau \alpha \beta \lambda \eta \tau \dot{\gamma} \gamma \varrho \alpha ́ \phi \varepsilon \tau \alpha \iota$

$$
y=f(\mathbf{x})+\varepsilon, \quad \varepsilon=v+w-u .
$$















 $\pi \omega ́ \lambda \eta \sigma \eta \varsigma ~ \varepsilon v o ́ \varsigma ~ \alpha к ı v \eta ́ \tau о v, ~ к о к) . ~ ' Н ~ \mu \pi о \varrho \varepsilon i ́ ~ v \alpha ~ \varepsilon к ф \varrho \alpha ́ \zeta о и v ~ \tau ı \varsigma ~ \alpha v \tau i ́ \theta \varepsilon \tau \varepsilon \varsigma ~ \tau \alpha ́ \sigma \varepsilon ı \varsigma ~ \pi о ט ~$









## MEPOг A : Е $\pi \iota \sigma \kappa o ́ \pi \eta \sigma \eta \kappa \alpha \iota A v \alpha ́ \lambda v \sigma \eta$

## КЕФ. 1: Елıбко́лךбך Вı $\beta \lambda \iota \gamma \varrho \alpha \phi i ́ \alpha \varsigma$.










 $\delta \iota \alpha \pi \varrho \alpha \gamma \mu \alpha ́ \tau \varepsilon v \sigma \eta \varsigma ~ " N a s h " \sigma \varepsilon \sigma \chi \varepsilon ́ \sigma \eta ~ \mu \varepsilon ~ \tau о v ~ \mu ı \sigma Ө o ́, ~ o ́ \pi о v ~ o ı ~ \varepsilon \pi \iota \delta \varrho \alpha ́ \sigma \varepsilon ı \varsigma ~ \tau \eta \varsigma ~ \varepsilon \lambda \lambda \varepsilon ı \pi о и ́ \varsigma ~$ $\pi \lambda \eta \varrho о ф о ́ \varrho \eta \sigma \eta \varsigma ~ \sigma v \nu v \pi \alpha ́ \varrho \chi о \cup v ~ \mu \varepsilon$ тıऽ $\varepsilon \pi \iota \delta \varrho \alpha ́ \sigma \varepsilon \iota \varsigma ~ \tau \eta \varsigma ~ \sigma \chi \varepsilon \tau \iota \kappa \eta ́ \varsigma ~ \delta \iota \alpha \pi \varrho \alpha \gamma \mu \alpha \tau \varepsilon v \tau \iota \kappa \eta ́ \varsigma$



 $\tau о \mu \varepsilon ́ \alpha$ ( $\chi \varrho \eta \mu \alpha \tau \iota \sigma \tau \eta ́ \varrho \iota \alpha \alpha \xi \iota \omega \prime \nu \kappa \alpha \iota \sigma v \nu \alpha \lambda \lambda \alpha ́ \gamma \mu \alpha \tau \circ \varsigma)$.
















 $\kappa \alpha \tau \alpha v о \mu \eta ́ \alpha v \alpha ́ \pi \alpha \varrho \alpha \tau \eta \varrho \emptyset \sigma \eta$.


 тццŋ́s / double price frontier").

## 





К $\alpha \tau \alpha \varrho \chi \alpha ́ \varsigma ~ \varepsilon \xi \varepsilon \tau \alpha ́ \zeta \varepsilon \tau \alpha \iota ~ \tau о ~ \delta о \mu ı к o ́ ~ v \pi o ́ \delta \varepsilon є \gamma \mu \alpha ~ \tau \omega \nu ~ P o l a c h e k ~ \& ~ Y o o n ~(1987) . ~ T o ~$








 عívaı $\eta \mu \varepsilon i ́ \omega \sigma \eta$ тov $\varepsilon \pi \iota \pi \varepsilon ́ \delta o v ~ \alpha \pi \alpha \sigma \chi$ о́ $\eta \sigma \eta \varsigma ~ \iota \sigma о \varrho \varrho о \pi i ́ \alpha \varsigma ~ \sigma \varepsilon ~ \sigma \chi \varepsilon ́ \sigma \eta ~ \mu \varepsilon \mu \iota \alpha ~ \kappa \alpha \tau \alpha ́ \sigma \tau \alpha \sigma \eta ~$




Aко入oú $\theta \omega \varsigma ~ \varepsilon \xi \varepsilon \tau \alpha ́ \zeta \varepsilon \tau \alpha \iota ~ \tau о ~ v \pi o ́ \delta \varepsilon เ \gamma \mu \alpha ~ \tau \omega v ~ G a y n o r ~ \& ~ P o l a c h e k ~(1994) . ~ T o ~ v \pi o ́ \delta \varepsilon ı \gamma \mu \alpha ~$





 $\pi \alpha \varrho \alpha \gamma o ́ v \tau \omega v$.





 $\pi \omega \dot{\lambda} \eta \sigma \eta \varsigma$.
 Kumbhakar \& Parmeter (2009). $\Delta \iota \alpha \pi \iota \sigma \tau \omega ́ v \varepsilon \tau \alpha \iota ~ o ́ \tau ı ~ \tau о ~ v \pi o ́ \delta \varepsilon \imath \gamma \mu \alpha ~ \chi \omega \lambda \alpha i ́ v \varepsilon ı ~ \alpha v \alpha ф о ю ı к \alpha ́ ~ \mu \varepsilon ~$








## MEPO $\Sigma$ B : Oıкоvo $\mu \varepsilon \tau \varrho i ́ \alpha \kappa \alpha \iota \Sigma \tau \alpha \tau \iota \sigma \tau \iota \kappa \eta ́$

## 











$$
\varepsilon=v+w-u, \quad v \sim N\left(0, \sigma_{v}^{2}\right), w \sim \operatorname{Exp}\left(\sigma_{w}\right), \quad u \sim \operatorname{Exp}\left(\sigma_{u}\right) .
$$



$$
f_{\varepsilon}(\varepsilon)=\frac{\exp \left\{a_{1}\right\} \Phi\left(b_{1}\right)+\exp \left\{a_{2}\right\} \Phi\left(b_{2}\right)}{\sigma_{w}+\sigma_{u}} .
$$



$$
\begin{aligned}
& F_{\varepsilon}(\varepsilon)=\Phi\left(\frac{\varepsilon}{\sigma_{v}}\right)+\frac{\sigma_{u}}{\sigma_{w}+\sigma_{u}} \exp \left\{a_{1}\right\} \Phi\left(b_{1}\right)-\frac{\sigma_{w}}{\sigma_{w}+\sigma_{u}} \exp \left\{a_{2}\right\} \Phi\left(b_{2}\right), \\
& a_{1}=\frac{\sigma_{v}^{2}}{2 \sigma_{u}^{2}}+\frac{\varepsilon}{\sigma_{u}}, \quad b_{1}=-\left(\frac{\varepsilon}{\sigma_{v}}+\frac{\sigma_{v}}{\sigma_{u}}\right), \quad a_{2}=\frac{\sigma_{v}^{2}}{2 \sigma_{w}^{2}}-\frac{\varepsilon}{\sigma_{w}}, \quad b_{2}=\frac{\varepsilon}{\sigma_{v}}-\frac{\sigma_{v}}{\sigma_{w}},
\end{aligned}
$$






$$
E\left(e^{w_{i}} e^{-u_{i}} \mid \varepsilon_{i}\right)=\frac{\exp \left\{\left(1+\sigma_{u}\right)\left(a_{1 i}+\frac{\sigma_{v}^{2}}{2 \sigma_{u}}\right)\right\} \Phi\left(b_{1 i}-\sigma_{v}\right)+\exp \left\{\left(1-\sigma_{w}\right)\left(a_{2 i}-\frac{\sigma_{v}^{2}}{2 \sigma_{w}}\right)\right\} \Phi\left(b_{2 i}+\sigma_{v}\right)}{\exp \left\{a_{1 i}\right\} \Phi\left(b_{1 i}\right)+\exp \left\{a_{2 i}\right\} \Phi\left(b_{2 i}\right)},
$$

 $\varepsilon v \tau \alpha ́ \sigma \sigma \varepsilon \tau \alpha \iota ~ \sigma \tau \eta \nu$ катпүоюí $\alpha \omega \nu$ " $\mu \varepsilon ́ \tau \varrho \omega \nu$ JLMS" ( $\beta \lambda$. Jondrow, Lovell, Materov \& Schmidt 1982).



$$
\operatorname{Pr}\left(w_{i}>u_{i} \mid \varepsilon_{i}\right)=\frac{\exp \left\{a_{2 i}\right\} \Phi\left(b_{2 i}\right)}{\exp \left\{a_{1 i}\right\} \Phi\left(b_{1 i}\right)+\exp \left\{a_{2 i}\right\} \Phi\left(b_{2 i}\right)} .
$$




$$
\varepsilon=v+w-u, \quad v \sim N\left(0, \sigma_{v}^{2}\right), w \sim H N\left(\sigma_{w}\right), \quad u \sim H N\left(\sigma_{u}\right),
$$



$$
\begin{aligned}
& f_{\varepsilon}\left(\varepsilon_{i}\right)=\frac{4}{s} \phi\left(\varepsilon_{i} / s\right)\left[\Phi_{2}\left(\frac{\varepsilon_{i}}{\omega_{1}}, 0 ; \rho=\frac{\lambda_{1}}{\sqrt{1+\lambda_{1}^{2}}}\right)-\Phi_{2}\left(\frac{\varepsilon_{i}}{\omega_{2}}, 0 ; \rho=\frac{-\lambda_{2}}{\sqrt{1+\lambda_{2}^{2}}}\right)\right], \\
& \theta_{1} \equiv \frac{\sigma_{w}}{\sigma_{v}}, \theta_{2} \equiv \frac{\sigma_{u}}{\sigma_{v}}, s \equiv \sqrt{\sigma_{v}^{2}+\sigma_{w}^{2}+\sigma_{u}^{2}}=\sigma_{v} \sqrt{1+\theta_{1}^{2}+\theta_{2}^{2}},
\end{aligned}
$$

$$
\omega_{1} \equiv \frac{s \sqrt{1+\theta_{2}^{2}}}{\theta_{1}}, \omega_{2} \equiv \frac{s \sqrt{1+\theta_{1}^{2}}}{\theta_{2}}, \quad \lambda_{1} \equiv \frac{\theta_{2}}{\theta_{1}} \sqrt{1+\theta_{1}^{2}+\theta_{2}^{2}}, \quad \lambda_{2} \equiv \frac{\theta_{1}}{\theta_{2}} \sqrt{1+\theta_{1}^{2}+\theta_{2}^{2}} .
$$


 К $\alpha$ vovıкŋ́s кат $\alpha$ vони́ร.
 ó@ $\omega v \alpha v \alpha ́ \pi \alpha \varrho \alpha \tau \eta ́ \varrho \eta \sigma \eta$.





 $\varepsilon \kappa \tau \mu \eta \tau \eta \dot{\varrho}$ @от $\omega$ v.



 $\tau \omega v \mu \varepsilon ́ \tau \rho \omega \nu$ JLMS.












 (cumulants), $\alpha v \tau i ́ \tau \omega \nu \sigma v v \eta \dot{\eta} \theta \omega \vee$ @o $\tau \omega \nu$ (moments).


 $\alpha v \tau 0 u ́ s ~ " k a p a-s t a t i s t i c s " ~ \gamma \iota \alpha ~ v \alpha ~ \delta \iota \alpha ф о \varrho о \pi о ь o u ́ v \tau \alpha \iota ~ \alpha \pi o ́ ~ \tau \alpha ~ " k-s t a t i s t i c s " ~(F i s h e r ~ 1930) ~ \pi o v ~$







$$
\left\{\begin{array}{l}
\hat{h}_{n, 2} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(c_{2}\right)^{-1} \hat{\varepsilon}_{i, O L S}^{2}-\left(\kappa_{2}(v)+\kappa_{2}(w)+\kappa_{2}(u)\right)\right] \\
\hat{h}_{n, 3} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(c_{3}\right)^{-1} \hat{\varepsilon}_{i, O L S}^{3}-\left(\kappa_{3}(w)-\kappa_{3}(u)\right)\right] \\
\hat{h}_{n, 4} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(C_{4 a} \hat{\varepsilon}_{i, O L S}^{4}-\frac{C_{4 b}}{(n-1)} \hat{\varepsilon}_{i, O L S}^{2} \sum_{\xi \neq i} \hat{\varepsilon}_{\xi, O L S}^{2}\right)-\left(\kappa_{4}(v)+\kappa_{4}(w)+\kappa_{4}(u)\right)\right] \\
\hat{h}_{n, 5} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(C_{5 a} \hat{\varepsilon}_{i, O L S}^{5}-\frac{C_{5 b}}{(n-1)} \hat{\varepsilon}_{i, O L S}^{3} \sum_{\xi \neq i} \hat{\varepsilon}_{\xi, O L S}^{2}\right)-\left(\kappa_{5}(w)-\kappa_{5}(u)\right)\right]
\end{array}\right.
$$


 $\tau \varepsilon \tau \varrho \alpha \gamma \omega v \omega \nu$.



$$
\hat{\mathbf{q}}: \hat{\mathbf{h}}_{N}(\hat{\mathbf{q}})=0 .
$$










```
\(v \sim N\left(0, \sigma_{v}^{2}\right), w \sim \operatorname{Gamma}(k, \theta), \quad u \sim \operatorname{Exp}\left(\sigma_{u}\right)\),
( \(\pi \alpha \varrho \alpha \lambda \lambda \alpha \gamma \eta \dot{\text { " "Г }} \mu \mu \alpha\)-ЕкӨєтєки́")
```

$\kappa \alpha \iota$
$v \sim N\left(0, \sigma_{v}^{2}\right), w \sim \operatorname{Exp}\left(\sigma_{w}\right), \quad u \sim \operatorname{Gamma}(k, \theta)$,


 ЕкӨєтьки́" غ́ $\chi о \cup \mu \varepsilon$

$$
f_{\varepsilon}(\varepsilon)=\frac{\sigma_{u}^{k-1}}{\left(\sigma_{u}+\theta\right)^{k}}\left[\exp \left\{\frac{\varepsilon}{\sigma_{u}}+\frac{\sigma_{v}^{2}}{2 \sigma_{u}^{2}}\right\}-\int_{0}^{\infty} e^{z / \sigma_{u}} \frac{1}{\sigma_{v}} \phi\left(\frac{\varepsilon-z}{\sigma_{v}}\right) F_{G}(z ; k, \delta) d z\right] .
$$



$f_{\varepsilon}(\varepsilon)=\frac{\sigma_{w}^{k-1}}{\left(\sigma_{w}+\theta\right)^{k}}\left[\exp \left\{-\frac{\varepsilon}{\sigma_{w}}+\frac{\sigma_{v}^{2}}{2 \sigma_{w}^{2}}\right\}-\int_{0}^{\infty} e^{z / \sigma_{w}} \frac{1}{\sigma_{v}} \phi\left(\frac{\varepsilon+z}{\sigma_{v}}\right) F_{G}(z ; k, \delta) d z\right], \quad \delta \equiv \sigma_{w} \theta /\left(\sigma_{w}+\theta\right)$
 vтокєф $\alpha \lambda \alpha \iota$.























 $\chi \supseteq \varnothing \sigma \mu$ отоเєít $\alpha$ เ.





 $\mu \varepsilon \tau \alpha \beta \lambda \eta \tau \eta \dot{w}$,

$$
f_{w}(w)=\frac{2}{\theta_{w}} \exp \left\{-w / \theta_{w}\right\}\left(1-\exp \left\{-w / \theta_{w}\right\}\right), \quad \theta_{w}>0, w \geq 0
$$

 $\alpha v \tau i ́ \sigma \tau о \chi \alpha \gamma \iota \alpha$ тоv ó@o $u$. К $\alpha$ vovi $\alpha \varsigma ~ \tau \eta \nu v \pi o ́ \theta \varepsilon \sigma \eta, \gamma \iota \alpha \quad \varepsilon=v+w-u$,

$$
v \sim N\left(0, \sigma_{v}^{2}\right), w \sim G E\left(2, \theta_{w}, 0\right), u \sim G E\left(2, \theta_{u}, 0\right),
$$



$$
\begin{aligned}
f_{\varepsilon}(\varepsilon)=\frac{2}{\theta_{w}+\theta_{u}}[ & \frac{2 \theta_{u} \exp \left\{a_{u}\right\} \Phi\left(b_{u}\right)}{\theta_{w}+2 \theta_{u}}-\frac{\theta_{u} \exp \left\{2 a_{u}+\left(\sigma_{v} / \theta_{u}\right)^{2}\right\} \Phi\left(b_{u}-\sigma_{v} / \theta_{u}\right)}{2 \theta_{w}+\theta_{u}} \\
& \left.+\frac{2 \theta_{w} \exp \left\{a_{w}\right\} \Phi\left(b_{w}\right)}{2 \theta_{w}+\theta_{u}}-\frac{\theta_{w} \exp \left\{2 a_{w}+\left(\sigma_{v} / \theta_{w}\right)^{2}\right\} \Phi\left(b_{w}-\sigma_{v} / \theta_{w}\right)}{\theta_{w}+2 \theta_{u}}\right]
\end{aligned}
$$

$\kappa \alpha \iota$

$$
\begin{aligned}
& F_{\varepsilon}(\varepsilon)= \frac{2}{\theta_{w}+\theta_{u}}\left[\frac{2\left(\theta_{w}+\theta_{u}\right)^{3}+\theta_{w} \theta_{u}\left(\theta_{w}+\theta_{u}\right)}{2\left(\theta_{w}+2 \theta_{u}\right)\left(2 \theta_{w}+\theta_{u}\right)} \Phi\left(\frac{\varepsilon}{\sigma_{v}}\right)\right. \\
&+\frac{2 \theta_{u}^{2}}{\theta_{w}+2 \theta_{u}} \exp \left\{a_{u}\right\} \Phi\left(b_{u}\right)-\frac{\theta_{u}^{2}}{2\left(2 \theta_{w}+\theta_{u}\right)} \exp \left\{2 a_{u}+\frac{\sigma_{v}^{2}}{\theta_{u}^{2}}\right\} \Phi\left(b_{u}-\frac{\sigma_{v}}{\theta_{u}}\right) \\
&\left.\quad-\frac{2 \theta_{w}^{2}}{2 \theta_{w}+\theta_{u}} \exp \left\{a_{w}\right\} \Phi\left(b_{w}\right)+\frac{\theta_{w}^{2}}{2\left(\theta_{w}+2 \theta_{u}\right)} \exp \left\{2 a_{w}+\frac{\sigma_{v}^{2}}{\theta_{w}^{2}}\right\} \Phi\left(b_{w}-\frac{\sigma_{v}}{\theta_{w}}\right)\right], \\
& a_{u}=\frac{\varepsilon}{\theta_{u}}+\frac{\sigma_{v}^{2}}{2 \theta_{u}^{2}}, \quad b_{u}=-\left(\frac{\varepsilon}{\sigma_{v}}+\frac{\sigma_{v}}{\theta_{u}}\right), \quad a_{w}=\frac{\sigma_{v}^{2}}{2 \theta_{w}^{2}}-\frac{\varepsilon}{\theta_{w}}, \quad b_{w}=\frac{\varepsilon}{\sigma_{v}}-\frac{\sigma_{v}}{\theta_{w}} .
\end{aligned}
$$


 $\kappa \alpha \tau \alpha v о \mu \eta ́ \varsigma \tau \eta \varsigma \delta \iota \alpha \phi$ оQ $\alpha \varsigma z=w-u$,

$$
\begin{aligned}
& f_{z}(z)= \begin{cases}\frac{2 \theta_{u}}{\theta_{w}+\theta_{u}}\left[\frac{2 \exp \left\{z / \theta_{u}\right\}}{\theta_{w}+2 \theta_{u}}-\frac{\exp \left\{2 z / \theta_{u}\right\}}{2 \theta_{w}+\theta_{u}}\right] \quad z \leq 0 \\
\frac{2 \theta_{w}}{\theta_{w}+\theta_{u}}\left[\frac{2 \exp \left\{-z / \theta_{w}\right\}}{2 \theta_{w}+\theta_{u}}-\frac{\exp \left\{-2 z / \theta_{w}\right\}}{\theta_{w}+2 \theta_{u}}\right] & z>0\end{cases} \\
& F_{z}(z)= \begin{cases}\frac{4 \theta_{u}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{z / \theta_{u}\right\}}{\left(\theta_{w}+2 \theta_{u}\right)}-\frac{\theta_{u}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{2 z / \theta_{u}\right\}}{\left(2 \theta_{w}+\theta_{u}\right)} & z \leq 0 \\
1-\frac{4 \theta_{w}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{-z / \theta_{w}\right\}}{2 \theta_{w}+\theta_{u}}+\frac{\theta_{w}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{-2 z / \theta_{w}\right\}}{\left(\theta_{w}+2 \theta_{u}\right)} & z>0\end{cases}
\end{aligned}
$$

 $\tau \not \mu \eta$ (conditional mode) $\kappa \alpha \iota \sigma \tau \eta \delta \varepsilon \mu \sigma \varepsilon \nu \mu \varepsilon ́ v \eta \delta_{\alpha} \mu \varepsilon \sigma \sigma$ (conditional median), $\pi \alpha \varrho \alpha ́ \alpha \sigma \tau \eta$ $\delta \varepsilon \sigma \mu \varepsilon \cup \mu \varepsilon ́ v \eta \mu \alpha Ө \eta \mu \alpha \tau \iota \kappa \eta ́ \varepsilon \lambda \pi i ́ \delta \alpha$.

 $\mu$ е́т@ךбŋऽ.




## 






 O@ $\gamma \alpha \nu \kappa \kappa \kappa v \mu \varepsilon \tau \alpha \beta \lambda \eta \tau \omega \prime v$ (Instrumental variables).

 Exponential Extension" rov Freund (1961),

$$
f_{w u}(w, u)=\left\{\begin{array}{ll}
a b^{\prime} \exp \left\{-b^{\prime} u-\left(a+b-b^{\prime}\right) w\right\} & 0<w<u \\
a^{\prime} b \exp \left\{-a^{\prime} w-\left(a+b-a^{\prime}\right) u\right\} & 0<u<w
\end{array} \quad a, a^{\prime}, b, b^{\prime}>0\right.
$$





 $\sigma v \nu \delta \iota \alpha \kappa \cup ́ \mu \alpha v \sigma \eta$ عívaı $\mu \eta \delta$ д́v.



$$
\begin{aligned}
& f_{\varepsilon}(\varepsilon)=\frac{a}{a+b} b^{\prime} \exp \left\{\frac{1}{2} \sigma_{v}^{2} b^{\prime 2}\right\} \exp \left\{b^{\prime} \varepsilon\right\} \Phi\left(-\frac{\varepsilon}{\sigma_{v}}-\sigma_{v} b^{\prime}\right) \\
&+\frac{b}{a+b} a^{\prime} \exp \left\{\frac{1}{2} \sigma_{v}^{2} a^{\prime 2}\right\} \exp \left\{-a^{\prime} \varepsilon\right\} \Phi\left(\frac{\varepsilon}{\sigma_{v}}-\sigma_{v} a^{\prime}\right)
\end{aligned}
$$





$$
\begin{gathered}
f_{\varepsilon}(\varepsilon)=\sqrt{2 \pi} \phi\left(\varepsilon / \sigma_{v}\right)\left[m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)\right] \\
\omega_{2} \equiv \frac{\varepsilon}{\sigma_{v}}+b^{\prime} \sigma_{v}, \quad \omega_{3} \equiv \frac{\varepsilon}{\sigma_{v}}-a^{\prime} \sigma_{v},
\end{gathered}
$$

$\varepsilon v \omega ́ ~ \eta ~ \tau \alpha v \tau о \pi о \iota \eta ́ \sigma \mu \eta ~ \sigma u v \alpha ́ \varrho \tau \eta \sigma \eta ~ к \alpha \tau \alpha v о \mu \eta ́ \varsigma ~ \mu \pi о \varrho \varepsilon i ́ ~ v \alpha ~ \gamma \varrho \alpha ф \varepsilon ́ ́$

$$
F_{\varepsilon}(\varepsilon)=\Phi\left(\varepsilon / \sigma_{v}\right)+\sqrt{2 \pi} \cdot \phi\left(\varepsilon / \sigma_{v}\right)\left[m \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)-(1-m) \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)\right] .
$$





$$
\begin{aligned}
& E(w)=\frac{a^{\prime}+b}{a^{\prime}(a+b)}=\frac{1}{a+b}+\frac{1-m}{a^{\prime}} \quad, \quad \operatorname{Var}(w)=\frac{a^{\prime 2}+2 a b+b^{2}}{a^{\prime 2}(a+b)^{2}}=\frac{1}{(a+b)^{2}}+\frac{1-m^{2}}{a^{\prime 2}}, \\
& E(u)=\frac{b^{\prime}+a}{b^{\prime}(a+b)}=\frac{1}{a+b}+\frac{m}{b^{\prime}} \quad, \quad \operatorname{Var}(u)=\frac{b^{\prime 2}+2 a b+a^{2}}{b^{\prime 2}(a+b)^{2}}=\frac{1}{(a+b)^{2}}+\frac{m(2-m)}{b^{\prime 2}},
\end{aligned}
$$

$$
\begin{aligned}
& \operatorname{Cov}(w, u)=\frac{a^{\prime} b^{\prime}-a b}{a^{\prime} b^{\prime}(a+b)^{2}}=\frac{1}{(a+b)^{2}}-\frac{m(1-m)}{a^{\prime} b^{\prime}}, \\
& \rho=\frac{a^{\prime} b^{\prime}-a b}{\sqrt{\left(a^{\prime 2}+2 a b+b^{2}\right)\left(b^{\prime 2}+2 a b+a^{2}\right)}} .
\end{aligned}
$$

 ótov $\pi$. $\chi$. غ́ $\chi o \cup \mu \varepsilon$

$$
\begin{aligned}
& E(w \mid \varepsilon)-\frac{1}{a+b}=\frac{\sigma_{v}(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\}\left[\omega_{3} \Phi\left(\omega_{3}\right)+\phi\left(\omega_{3}\right)\right]}{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)}, \\
& E(u \mid \varepsilon)-\frac{1}{a+b}=\frac{\sigma_{v} m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\}\left[\phi\left(\omega_{2}\right)-\omega_{2} \Phi\left(-\omega_{2}\right)\right]}{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)} .
\end{aligned}
$$



 $\alpha v \alpha ́ \lambda v \sigma \eta \varsigma ~ \sigma \tau о \chi \alpha \sigma \tau \iota \kappa o v ́ ~ \sigma u v o ́ \rho o v . ~$

 $\pi \nu \kappa \nu o ́ \tau \eta \tau \alpha \varsigma \pi \iota \theta \alpha \nu$ о́т $\eta \tau \alpha \varsigma$

$$
f_{Z}(z)= \begin{cases}m b^{\prime} \exp \left\{b^{\prime} z\right\} & z<0 \\ (1-m) a^{\prime} \exp \left\{-a^{\prime} z\right\} & z \geq 0\end{cases}
$$



$$
F_{z}(z)= \begin{cases}m \exp \left\{b^{\prime} z\right\} & z<0 \\ 1-(1-m) \exp \left\{-a^{\prime} z\right\} & z \geq 0\end{cases}
$$



 $\mu \varepsilon \tau \alpha \beta \lambda \eta \tau \eta ́ \varsigma$.
 $\tau \eta \sigma \chi \varepsilon ́ \sigma \eta$

$$
\operatorname{sign}\{\operatorname{Cov}(w, u)\}=\operatorname{sign}\{\operatorname{Cov}(E(w \mid \varepsilon), E(u \mid \varepsilon))\}=\operatorname{sign}\left\{\operatorname{Cov}\left(E(w \mid \varepsilon)-\frac{1}{a+b}, E(u \mid \varepsilon)-\frac{1}{a+b}\right)\right\}
$$



 E入é $\gamma \chi$ о⿱

$$
\hat{q}=\frac{\left((1-\hat{m}) \hat{a}^{\prime}-\hat{m} \hat{b}^{\prime}\right)^{2}}{\hat{\operatorname{Var}}(T)} \xrightarrow[\mathrm{H}_{0}]{d} \chi_{1}^{2}
$$

о́тоv

$$
\begin{gathered}
\operatorname{Var}(T)=(1-m)^{2} \operatorname{Var}\left(\hat{a}^{\prime}\right)+m^{2} \operatorname{Var}\left(\hat{b}^{\prime}\right)+\left(a^{\prime}+b^{\prime}\right)^{2} \operatorname{Var}(\hat{m})-2(1-m) m \operatorname{Cov}\left(\hat{a}^{\prime}, \hat{b}^{\prime}\right) \\
-2(1-m)\left(a^{\prime}+b^{\prime}\right) \operatorname{Cov}\left(\hat{a}^{\prime}, \hat{m}\right)+2\left(a^{\prime}+b^{\prime}\right) m \operatorname{Cov}\left(\hat{m}, \hat{b}^{\prime}\right) .
\end{gathered}
$$


 $\varepsilon к \tau \iota \mu \tau \eta$.
 Kó $\pi о \cup \lambda \alpha$ (Copula) $\gamma \iota \alpha$ тךv $\alpha \nu \tau \iota \mu \varepsilon \tau \omega ́ \pi \iota \sigma \eta$ тоv $\phi \alpha \iota v o \mu \varepsilon ́ v o v ~ \tau \eta \varsigma ~ \varepsilon v \delta o \gamma \varepsilon ́ v \varepsilon \iota \alpha \varsigma ~ \tau \omega v$
 (2015), о́ $\mu \omega \varsigma ~ \pi \varepsilon \varrho \iota ү \varrho \alpha ́ \phi \varepsilon \iota ~ \varepsilon \pi \iota \pi \varrho о \sigma \theta \varepsilon ́ \tau \omega \varsigma ~ \alpha v \alpha \lambda v \tau \iota \kappa \alpha ́ ~ o ́ \lambda \alpha ~ \tau \alpha ~ Ө \varepsilon ́ \mu \alpha \tau \alpha ~ \pi о v ~ \mu \pi о \varrho \varepsilon i ́ ~ v \alpha ~$




Афои́ $\pi \alpha \varrho о v \sigma \iota \alpha \sigma \tau о$ и́v $\pi \varepsilon \varrho \iota \lambda \eta \pi \tau \iota \kappa \alpha ́ \alpha$ оı $\beta \alpha \sigma \iota \kappa \varepsilon ́ \varsigma ~ \alpha \varrho \chi \varepsilon ́ \varsigma ~ \tau \eta \varsigma ~ \theta \varepsilon \omega \varrho i ́ \alpha \varsigma ~ \tau \omega \nu ~ К о ́ \pi о \cup \lambda \alpha$, $\alpha \kappa о \lambda$ ои́ $\Theta \omega \varsigma ~ \pi \alpha \varrho о v \sigma \iota \alpha ́ \zeta о v \tau \alpha \iota$ oı $\pi \varrho \alpha \kappa \tau \iota \kappa \varepsilon ́ \varsigma ~ \mu \varepsilon ́ \theta$ oठoı $\varepsilon \phi \alpha \varrho \mu о \gamma \eta ́ \varsigma ~ \tau о \cup \varsigma ~ \sigma \varepsilon ~ \varepsilon ́ v \alpha ~ v \pi o ́ \delta \varepsilon \iota \gamma \mu \alpha$ $\pi \alpha \lambda \iota v \delta \varrho о ́ \mu \eta \sigma \eta \varsigma$,

$$
y_{i}=\mathbf{x}_{i}^{\prime} \beta+\varepsilon_{i}, \quad \varepsilon_{i}=v_{i}+w_{i}-u_{i}, \quad i=1, \ldots, n .
$$




$$
\ell_{i}=\ln c\left[F_{1}\left(x_{1 i}\right), \ldots, F_{m}\left(x_{m i}\right), F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta\right)\right]+\ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta\right),
$$






- $\Sigma \tau \eta v$ Kó $\tau о \cup \lambda \alpha \delta \varepsilon v \varepsilon \mu \phi \alpha v i ́ \zeta o v \tau \alpha \iota \tau v \chi o ́ v v \tau \varepsilon \tau \varepsilon \varrho \mu \iota v \iota \sigma \tau \iota \kappa \varepsilon ́ \varsigma \mu \varepsilon \tau \alpha \beta \lambda \eta \tau \varepsilon ́ \varsigma$
- $\Sigma \tau \eta v$ Kó $\pi o v \lambda \alpha$ $\varepsilon \mu \phi \alpha v i ́ \zeta o v \tau \alpha \iota \mu o ́ v o$ ol $\tau v \chi \alpha i ́ \varepsilon \varsigma ~ \mu \varepsilon \tau \alpha \beta \lambda \eta \tau \varepsilon ́ \varsigma ~ \pi o v ~ \theta \varepsilon \omega \varrho o u ́ v \tau \alpha \iota$ evסoүعveís


 $\delta \varepsilon \nu \varepsilon \mu \phi \alpha v i ́ \zeta о \nu \tau \alpha \iota \sigma \tau \eta \nu$ Ко́ $\tau о \cup \lambda \alpha$.



$$
\breve{x}_{j i} \equiv \hat{F}_{j}\left(x_{j i}\right)=\frac{1}{n+1} \sum_{k=1}^{n} I\left\{x_{j k} \leq x_{j i}\right\}, \quad i=1, \ldots, n, \quad j=1, \ldots, m,
$$



$$
\breve{x}_{j i}=\hat{F}_{j}\left(x_{j i}\right)=\frac{1}{n+1} \sum_{k=1}^{n} I\left\{x_{j k}<x_{j i}\right\}+\frac{1}{n+1} \sum_{k=1}^{n} \frac{1}{2} I\left\{x_{j k}=x_{j i}\right\} .
$$



$$
\begin{aligned}
& \ln c_{i}^{G}=-\frac{1}{2} \ln \operatorname{det}(\mathrm{R})-\frac{1}{2} \mathbf{q}_{i}^{\prime}\left(\mathrm{R}^{-1}-I_{m+1}\right) \mathbf{q}_{i} \\
& \mathbf{q}_{i}=\left(\Phi^{-1}\left(\breve{x}_{1 i}\right), \ldots, \Phi^{-1}\left(\breve{x}_{m i}\right), \Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right)^{\prime} .
\end{aligned}
$$



$$
\mathrm{R}=\left[\begin{array}{cccc}
1 & \rho_{12} & \cdots & \rho_{1 \varepsilon} \\
\rho_{12} & 1 & & \vdots \\
\vdots & & 1 & \rho_{m \varepsilon} \\
\rho_{1 \varepsilon} & \cdots & \rho_{m \varepsilon} & 1
\end{array}\right]
$$

О $\mu \varepsilon \tau \alpha \sigma \chi \eta \mu \alpha \tau \iota \sigma \mu o ́ s \quad \Phi^{-1}\left(\bar{x}_{j i}\right) \quad \beta \alpha \sigma i \zeta \varepsilon \tau \alpha \iota \quad \sigma \tau \sigma \nu \quad$ Мعт $\alpha \sigma \chi \eta \mu \alpha \tau \iota \sigma \mu o ́ \quad \tau о v$










$$
\begin{aligned}
\ln \breve{L}=-\frac{n}{2} \ln \operatorname{det}(\breve{\mathrm{R}})- & \frac{1}{2} \sum_{i=1}^{n} \mathbf{q}_{i}^{\prime} \breve{\mathrm{R}}^{-1} \mathbf{q}_{i} \\
& +\frac{1}{2} \sum_{i=1}^{n}\left[\Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right]^{2}+\sum_{i=1}^{n} \ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right),
\end{aligned}
$$

$$
\breve{\mathrm{R}}=\left[\begin{array}{ccccc}
1 & \hat{\rho}_{12} & \cdots & \hat{\rho}_{1 m} & \rho_{1 \varepsilon} \\
\hat{\rho}_{12} & 1 & \cdots & \hat{\rho}_{2 m} & \rho_{2 \varepsilon} \\
\vdots & \vdots & \ddots & & \vdots \\
\hat{\rho}_{1 m} & \vdots & & \ddots & \rho_{m \varepsilon} \\
\rho_{1 \varepsilon} & \cdots & \cdots & \rho_{m \varepsilon} & 1
\end{array}\right]
$$




 $\sigma \chi \varepsilon ́ \sigma \varepsilon เ \varsigma$.







 $\varepsilon v \delta o \gamma \varepsilon ́ v \varepsilon \iota \alpha \varsigma$.


 p.99).



5) $\Upsilon \pi \alpha ́ \varrho \chi о \cup \nu \alpha \pi о \tau \varepsilon \lambda \varepsilon ́ \sigma \mu \alpha \tau \alpha \pi \varrho о \sigma о \mu о \iota \omega \sigma \varepsilon \omega \nu$ Monte Carlo $\sigma ט ́ \mu \phi \omega v \alpha \mu \varepsilon \tau \alpha$ отоí $\alpha \eta$




 $\varepsilon \lambda \varepsilon ́ \gamma \chi о \cup \mu \varepsilon$ к $\alpha \tau \alpha ́$ тóбo oı $\mu \varepsilon \tau \alpha \sigma \chi \eta \mu \alpha \tau \iota \sigma \mu \varepsilon ́ v \varepsilon \varsigma ~ \varepsilon \pi \varepsilon \xi \eta \gamma \eta \mu \alpha \tau \iota \kappa \varepsilon ́ \varsigma ~ \mu \varepsilon \tau \alpha \beta \lambda \eta \tau \varepsilon ́ \varsigma ~ \alpha \kappa о \lambda о \cup \theta$ оóv
 $\gamma \nu \omega \sigma \tau \varepsilon ́ \varsigma ~ \mu \varepsilon Ө$ обо入о $\gamma і ́ \varepsilon \varsigma ~ \sigma \tau \eta ~ \beta ı \beta \lambda \iota о \gamma \varrho \alpha \phi \dot{\alpha} \alpha$.









 $\mu \varepsilon \lambda \varepsilon ́ \tau \varepsilon \varsigma$.

## 

## КЕФ. 5: 'Eva véo 2TSF vтóठєı $\gamma \mu \alpha$ סıцع@ov́s $\delta \iota \alpha \pi \varrho \alpha \gamma \mu \alpha ́ \tau \varepsilon v \sigma \eta \varsigma ~ N a s h ~$  $\pi \varrho о \sigma \delta \iota о \varrho \iota \sigma \mu$ о́ тоv $\mu \iota \sigma \theta$ ои́ vто́ к $\alpha \theta \varepsilon \sigma \tau \omega ́ \varsigma \alpha \beta \varepsilon ́ \beta \alpha ı \eta \varsigma ~ \pi \alpha \varrho \alpha \gamma \omega \gamma \iota \kappa о ́ \tau \eta \tau \alpha \varsigma$.

К $\alpha \tau \alpha ́ ~ \tau \eta v ~ \alpha v \alpha ́ \lambda v \sigma \eta ~ \tau \omega v ~ \delta о \mu \iota \kappa \omega ́ v ~ \theta \varepsilon \mu \varepsilon \lambda i ́ \omega v ~ \tau о v ~ v \tau о д \varepsilon ́ ́ \gamma \mu \alpha \tau о \varsigma ~ 2 T S F ~ \sigma \tau о ~ к \varepsilon ф \alpha ́ \lambda \alpha เ о ~ 2, ~$







 $\sigma \chi \varepsilon ́ \sigma \eta ~ \pi \varrho о \sigma \delta \iota \varrho \varrho \iota \sigma \mu \circ$ и́ тov $\mu \iota \sigma \theta$ ои́:

$$
\omega^{*}=E\left(p \mid I_{f} \cap I_{e}\right)+v+\eta g-(1-\eta) d
$$

о́тои
$\omega^{*}$ : o $\mu \iota \sigma$ Ós $\pi 0 v \tau \varepsilon \lambda \iota \kappa \alpha ́ \sigma \nu \mu \phi \omega v \varepsilon i ́ t \alpha \iota$

 غ́v $\alpha \varrho \xi ̌ \eta \tau \eta \varsigma \delta \iota \alpha \pi \varrho \alpha \gamma \mu \alpha ́ \tau \varepsilon v \sigma \eta \varsigma$


 $\chi \propto \varrho \alpha \kappa \tau \eta \varrho เ \sigma \tau \iota \kappa \alpha ́ ~ \tau \eta \varsigma ~ \varepsilon \pi \tau \chi \varepsilon i ́ \varrho \eta \sigma \eta \varsigma ~ \kappa \lambda \pi$.
 $\delta \iota \pi \varrho \alpha \gamma \mu \alpha ́ \tau \varepsilon v \sigma \eta$.


 $\pi \varepsilon ́ \varrho \alpha \kappa \alpha \iota ~ \pi \alpha ́ v \omega \tau \eta \varsigma E\left(p \mid I_{f} \cap I_{e}\right)$.



 vтодєí $\alpha$ тоऽ 2TSF, $\varepsilon v \omega ́ ~ \eta ~ \delta \varepsilon \sigma \mu \varepsilon v \mu \varepsilon ́ v \eta ~ \mu \alpha \theta \eta \mu \alpha \tau \iota \kappa \eta ́ ~ \varepsilon \lambda \pi i ́ \delta \alpha E\left(p \mid I_{f} \cap I_{e}\right) \varepsilon \kappa ф \varrho \alpha ́ \zeta \varepsilon \tau \alpha \iota \omega \varsigma$



 $\varepsilon \pi \varepsilon \zeta \dot{\eta} \gamma \eta \sigma \eta \varsigma$ tov $\mu$ ІбӨov́ $\alpha \lambda \alpha$ Mincer.



1) $\uparrow \uparrow \pi \alpha \varrho \xi \eta ~ \alpha \sigma ט ́ \mu \mu \varepsilon \tau \varrho \eta \varsigma ~ \kappa \alpha \iota ~ \varepsilon \tau \varepsilon \varrho о \gamma \varepsilon v o u ́ \varsigma ~ \pi \lambda \eta \varrho о ф о ́ \varrho \eta \sigma \eta \varsigma$,



 $\delta \iota \alpha \pi \varrho \alpha \gamma \mu \alpha ́ \tau \varepsilon \nu \sigma \eta$,









 $\varepsilon \pi \varepsilon \xi \eta \gamma \eta \mu \alpha \tau \iota \kappa \varepsilon ́ \varsigma \mu \varepsilon \tau \alpha \beta \lambda \eta \tau \varepsilon ́ \varsigma \kappa \alpha \iota \gamma \iota \alpha \tau \iota \varsigma \delta$ ט́o $\mu \varepsilon ́ \varrho \eta \tau \eta \varsigma \delta \iota \alpha \pi \varrho \alpha \gamma \mu \alpha ́ \tau \varepsilon v \sigma \eta \varsigma$.



 $\alpha v \tau i ́ \theta \varepsilon \tau \eta \varsigma$ фо@д́s.











 $\mu$ оvт $\alpha \dot{\lambda}$ ov Ко́ $\pi о \cup \lambda \alpha \pi \alpha \varrho \alpha ́ \lambda \lambda \eta \lambda \alpha$.











##  (management) $\sigma \tau \eta v \pi \alpha \varrho \alpha \gamma \omega \gamma \eta \dot{\eta} \kappa \alpha \iota \quad \gamma \iota \alpha$ то $\pi \varrho o ́ \beta \lambda \eta \mu \alpha$ тŋร $\lambda \alpha v \theta \alpha \sigma \mu \varepsilon ́ v \eta \varsigma$ $\alpha \sigma v \mu \mu \varepsilon \tau \varrho i ́ \alpha \varsigma($ ("wrong skewness problem") $\sigma \tau о \chi \alpha \sigma \tau \iota \kappa \eta ́ \alpha v \alpha ́ \lambda v \sigma \eta$ $\sigma v v o ́ \varrho o v$.


















4) $\Upsilon \pi \alpha ́ \varrho \chi \varepsilon \iota ~ к \alpha ́ \pi о ь о ~ " \tau \varepsilon \chi ข เ \kappa \alpha ́ ~ \alpha ́ \varrho เ \sigma \tau о ~ \varepsilon \pi i ́ \pi \varepsilon \delta o ~ m a n a g e m e n t " ; ~ М \pi о \varrho \varepsilon i ́ ~ \mu ı \alpha ~ " \alpha u ́ \xi \eta \sigma \eta " ~$
 $\pi \alpha \varrho \alpha \gamma o ́ \mu \varepsilon v o v ~ \pi \varrho o$ ö́vtos; Eívaı $\eta$ हाíठ@ $\alpha \sigma \eta$ tov management $\varepsilon \pi i ́ ~ \tau o v$ $\pi \alpha \varrho \alpha \gamma$ о́ $\mu \varepsilon$ vov $\pi \varrho о$ ö́vтоऽ $\alpha ́ v \omega$ ф@ $\alpha \gamma \mu \varepsilon ́ v \eta ;$

 $\mu \eta \delta \varepsilon v i ́ \sigma \varepsilon \iota \tau \eta v \pi \alpha \varrho \alpha \gamma \omega \gamma \eta ;$
 $\kappa \alpha \tau \alpha \sigma \kappa \varepsilon \cup \eta ́ ~ \varepsilon v o ́ \varsigma ~ " m a n a g e m e n t ~ s c o r e " ~ \beta \alpha \sigma \iota \sigma \mu \varepsilon ́ v o ~ \sigma \varepsilon ~ \pi \varrho о \sigma \omega \pi \iota \kappa \varepsilon ́ \varsigma ~ \sigma u v \varepsilon v \tau \varepsilon u ́ \xi \varepsilon \iota \varsigma ~ \mu \varepsilon ~$



 Bloom єктíцךбє то ко́бтоs тךऽ ка́Өє $\mu i ́ \alpha \varsigma ~ \pi \alpha \varrho \alpha \tau \eta ́ \varrho \eta \sigma \eta \varsigma ~ \sigma \varepsilon ~ \pi \varepsilon \varrho i ́ т о v ~ 500 ~ \delta о \lambda \lambda \alpha ́ \varrho \iota \alpha ~ Н П А . ~$
 бо $\lambda \lambda \alpha ́ \varrho \iota \alpha ~ Н П А) . ~$



 $\alpha v \mu \pi о \varrho \varepsilon i ́ ~ v \alpha ~ \mu \varepsilon \tau \varrho \eta \theta \varepsilon i ́ ~ \eta ~ \sigma v v \varepsilon เ \sigma ф о \varrho \alpha ́ ~ \tau \eta \varsigma, ~ \delta \varepsilon \delta о \mu \varepsilon ́ v o v ~ o ́ \tau \iota ~ \alpha v \tau \eta ́ ~ v \lambda о \pi о \iota \varepsilon i ́ t \alpha ı ~ \sigma \varepsilon ~$







 тоьотเкои́s, $\alpha \pi$ ó $\lambda v \tau 0 \cup \varsigma ~ \eta ́ ~ \sigma \chi \varepsilon \tau \iota \kappa o u ́ s . ~ \Sigma \tau \eta v ~ \pi \varrho \alpha \gamma \mu \alpha \tau \iota \kappa o ́ \tau \eta \tau \alpha ~ \alpha v \tau o ́ ~ \pi о v ~ \varepsilon к \tau \iota \mu \alpha ́ ~ \tau о ~$

 $h()$.
3) To v $\tau$ ó $\delta \varepsilon \iota \gamma \mu \alpha$ عívaı $\sigma \tau \alpha \tau \iota \kappa o ́, ~ \varepsilon \sigma \tau \iota \alpha ́ \zeta о v \tau \alpha \varsigma ~ \sigma \varepsilon ~ \delta \iota \alpha \sigma \tau \varrho \omega \mu \alpha \tau \iota \kappa \alpha ́ ~ \delta \varepsilon \delta о \mu \varepsilon ́ v \alpha ~ \kappa \alpha \iota ~ \kappa \alpha \tau \alpha ́ ~$


4) $\Upsilon \pi о \delta \varepsilon \downarrow \gamma \mu \alpha \tau о \pi о \iota \omega ́ v \tau \alpha \varsigma$ то management $\omega \varsigma \mu \varepsilon \tau \alpha \beta \lambda \eta \tau \eta$, то $v \pi o ́ \delta \varepsilon \downarrow \gamma \mu \alpha \sigma v \mu \beta \alpha \delta i ́ \zeta \varepsilon$


5) To management $\delta \varepsilon \nu \mu \pi$ о@ $\varepsilon$ í $v \alpha \mu \eta \delta \varepsilon v i ́ \sigma \varepsilon \iota ~ \tau \eta \nu \pi \alpha \varrho \alpha \gamma \omega \gamma \eta$. H $\varepsilon \lambda \alpha ́ \chi \iota \sigma \tau \eta \sigma v v \varepsilon \iota \sigma \phi \circ \varrho \alpha ́$



 об $\eta \gamma \eta \dot{\sigma} \sigma \iota \mu \iota \alpha \varepsilon \pi \iota \chi \varepsilon$ í@ $\eta \sigma \eta$ $\sigma \varepsilon \alpha \phi \alpha \nu \iota \sigma \mu o ́)$.


 عเб○ои́v.

Н $\sigma u v \alpha ́ \varrho \tau \eta \sigma \eta \pi \alpha \varrho \alpha \gamma \omega \gamma \eta ́ s ~ \varepsilon i ́ v \alpha \iota$

$$
Q=A F(\mathbf{x}) e^{h(m)}, \quad m \geq 0, \quad h(0)=0, \quad h^{\prime}(m)>0, \quad h^{\prime \prime}(m)<0 .
$$

ó ó





B $\alpha \sigma \iota \kappa \alpha ́ \alpha \pi о \tau \varepsilon \lambda \varepsilon ́ \sigma \mu \alpha \tau \alpha$ ع $\delta \omega$ عív $\alpha$ :





 оккоvoцía.

 Кєф $\alpha \lambda \alpha$ íov/Ее $\gamma \alpha \sigma$ б́ $\alpha$.








$$
Q=A F(\mathbf{x}) e^{h(m)} e^{v} e^{-u}, \quad E(v)=0, \quad u \geq 0
$$




 $\pi \varepsilon \varrho \prec \beta \alpha \lambda \lambda$ ov.







 $w \equiv h(m):$

$M c \equiv \frac{Q-A F(\mathbf{x}) e^{v} e^{-u}}{Q}=1-\frac{A F(\mathbf{x}) e^{v} e^{-u}}{A F(\mathbf{x}) e^{v+w-u}}=1-e^{-w}$.
 management $\sigma \tau \eta \nu \pi \alpha \varrho \alpha \gamma \omega \gamma \eta$, $\omega \varsigma \pi \alpha \varrho \alpha \gamma \omega \gamma \iota \kappa \eta$ єıఠ@oŋ́.


$$
\operatorname{ExIn} \equiv \frac{Q e^{u}-Q}{Q e^{u}}=1-e^{-u}
$$


$M s \equiv e^{w}$.

$E D \equiv e^{w} e^{-u}$.




 $\varepsilon \pi i ́ \pi \varepsilon \delta \alpha \alpha \pi о \varrho \varrho о ́ \phi \eta \sigma \eta \varsigma \tau \omega v \sigma \nu \mu \beta \alpha \tau \iota \kappa \omega ́ v \pi \alpha \varrho \alpha \gamma \omega \gamma \iota \omega \dot{v} \varepsilon \iota \sigma \varrho \circ \omega ้ v$.
 $\delta \varepsilon \delta о \mu \varepsilon ́ v \omega \nu \gamma \iota \alpha \tau \alpha$ отоі́ $\alpha$ v $\quad \alpha ́ \varrho \chi о \cup v \delta \delta \iota \theta \varepsilon ́ \sigma \not \mu \alpha \tau \alpha$ management scores $\tau \omega v$ Bloom \& Van





 $\Delta$ เоíкпбпऽ бто т@оїóv.








 $\sigma \tau \circ \chi \varepsilon i ́ \omega v$ к $\alpha \iota ~ v \pi о \lambda о \gamma \iota \sigma \mu \circ v ́ \tau \omega v$ BVRA scores.
 $\phi \alpha เ v o ́ \mu \varepsilon v o ~ \tau \eta \varsigma ~ " \lambda \alpha v \theta \alpha \sigma \mu \varepsilon ́ v \eta \varsigma ~ \alpha \sigma v \mu \mu \varepsilon \tau \varrho i ́ \alpha \varsigma " ~(w r o n g ~ s k e w n e s s) ~ \pi o v ~ \pi \alpha \varrho \alpha \tau \eta \varrho \varepsilon i ́ t \alpha \iota ~ \sigma \varepsilon ~$




 غ́ $\chi$ оuv $\Pi \varrho о \tau \alpha \theta \varepsilon$ í.

 $\sigma \tau 0$ ótı то management $\delta \varepsilon v \lambda \alpha \mu \beta \alpha ́ v \varepsilon \tau \alpha \iota ~ v \pi o ́ \psi \varepsilon \iota ~ \sigma \tau \alpha ~ v \pi о \delta \varepsilon i ́ \gamma \mu \alpha \tau \alpha ~ \alpha v \tau \alpha ́ . ~ К \alpha \tau \alpha ́ ~ \sigma v \nu \varepsilon ́ \pi \varepsilon ı \alpha$,














 $\pi \alpha \varrho \alpha \gamma \omega \gamma \iota \kappa$ ои́ $\mu \eta \chi \alpha \nu \iota \sigma \mu$ ои́.








## 




 кข@í $\omega \varsigma$ кє́́ $\mu \varepsilon v о$.

## (ТЕЛОГ ПЕРІЛНЧНГ)

## Introduction \& Overview

Stochastic frontier analysis (SFA) is one of the two main scientific approaches in measuring efficiency, or business performance. ${ }^{1}$ As Fried, Lovell \& Schmidt (2008) write in the preface of their book, "business performance" should be thought of broadly as "doing the right things right". ${ }^{2}$ Although "frontier analysis" originates back to the introduction of distance functions in economics by Shepherd (1953, 1970), the stochastic/econometric character of the field begun with two papers, Aigner, Lovell \& Schmidt (1977) and Meeusen \& van den Broeck (1977), that introduced for the first time a stochastic component into the functional specification.

Commenting on the fundamental concept of output-oriented technical efficiency (which is the concept of efficiency that applies to any mechanism of production irrespective of its financial goal) Kumbhakar \& Lovell (2000) write (p. 72): "The great virtue of stochastic production frontier models is that the impact on output of shocks due to variation in labor and machinery performance, vagaries of the weather, and just plain luck can at least in principle be separated from the contribution of variation in technical efficiency".

It is this ability to separate and disentangle forces and influences on observed outcomes of (mainly) economic activities that has been taken to another level by the two-tier stochastic frontier (2TSF) framework introduced by Polachek \& Yoon (1987): in its context, the phenomenon of competing forces on observed outcomes, forces that pull the outcome to opposing directions, can be measured and assessed even if we do not possess data on these influences. It is no accident that, as we shall see in detail in chapter 1, the model has been used to also analyze some surprising and certainly not-economic situations and research questions.

In abstract terms, assume that we have an outcome/dependent variable $y$ for which we postulate that it is a function $f(\cdot)$ of a vector of explanatory variables/regressors $\mathbf{x}$ on

[^0]which we have data. The dependent variable is further affected by a stochastic noise/disturbance/shock/error denoted $v$. Assume now that we know (or that we can adequately argue) that apart from $\mathbf{x}$ and $v$, there exist two forces that affect $y$ each in the opposite direction, but for which we possess no data. Denote the positive influence $w$ and the negative influence $u$. Since we have no data on them, it is natural to treat them both as one-sided positive random variables attaching a negative sign to the one representing the negative influence. Then the expression for the dependent variable becomes
$$
y=f(\mathbf{x})+\varepsilon, \quad \varepsilon=v+w-u .
$$

Cross-sectional SFA models in their various incarnations use either $u$ or $w$ and in that case $f(\mathbf{x})$ is interpreted as a frontier, an upper or lower boundary as the case may be. The 2TSF framework uses both, and it allows us to obtain information on $u$ and $w$ separately from the random disturbance $v$, as well as separately from each other. It thus achieves the goal of revealing, disentangling and quantitatively assessing these unobserved influences on the outcome $y$, enriching our understanding of the phenomenon under study.

It also transforms the representation of the frontier: in single-tier SF models, $f(\mathbf{x})$ is the deterministic frontier, and $f(\mathbf{x})+v$ its stochastic counterpart, while $w$ or $u$ (which one is present) measure the distance from it. In the 2TSF context there are two inherently stochastic frontiers, and $w$ or $u$ take part in determining the one while measuring the distance from the other: $f(\mathbf{x})+v+w$ represents the stochastic upper frontier (or the frontier of the "seller" in an economic exchange) while $u$ represents the distance from it. At the same time, $f(\mathbf{x})+v-u$ represents the stochastic lower frontier (or the frontier of the buyer), and here it is $w$ that measures the distance from the latter. The difference of the two frontiers equals $w+u$ and measures the length of the interval into which the stochastic outcome can lie.

It is evident that the above very general framework is applicable to a large number of circumstances: indicatively, the competing forces on the outcome can be informational deficiencies of the buyer and of the seller in an economic transaction that affect the realized price (which can be the wage of an employee, the fee of a doctor, or the price of a house); or agency conflicts inside a corporation; or over-bidding and under-bidding strategies in an
auction. All the above have been examined in the 2TSF literature, together with some more exotic ones. It should be clear that the two-tier stochastic frontier framework is not constrained to model and measure efficiency in resource use in production and economic activity. It is not just a special case of stochastic frontier analysis but a general modeling and estimation framework for any situation that possesses the structure described above.

After a 20-year period of rather sparse application, the 2TSF approach has started to increasingly attract the attention of researchers, and this PhD thesis aims to contribute to this momentum by offering new theoretical results that make the validity of the model more robust, construct new modeling and estimation tools that enhance the options available to the applied researcher, and elaborate on theoretical and empirical applications of the 2TSF model in new situations where competing unobserved forces impact the observed outcome.

The thesis is structured and written as a monograph rather than as a compilation of stand-alone research papers. Certainly the material here has aspirations to be published in peer-reviewed platforms, in which case the various chapters will be reworked in order to acquire an autonomous presence (and most likely the empirical applications will turn into empirical studies, possibly with different data samples). The thesis is organized in four parts:

Part A - Review and Analysis includes chapters 1 (Literature) and 2 (Structural Foundations). We have tried to make the literature review as comprehensive as possible, and indeed apparently exhaustive as of July 2018, since the number of papers dealing with the 2TSF framework is not large.

In chapter 2 we analyze, compare and contrast the structural foundations that have been proposed in the literature that lead to a 2TSF reduced-form model. This is the first time that such an analysis has been undertaken, and with it we attempt to shed light, clarify and strengthen the underlying mechanisms the lead to a 2TSF estimation framework.

Part B - Econometrics and Statistics presents tools to estimate 2TSF models, and includes chapter 3 ("Independence and Exogeneity") and chapter 4 ("Dependence and Endogeneity"): the material is grouped according to these two assumptions because they
determine which tools can be used in each case. In chapter 3 we assume that the three error components in a 2 TSF specification are jointly independent, and that the regressors are exogenous to the error term. For completeness, we start by presenting the benchmark Exponential 2TSF specification for the composite error term, while enhancing it with several results previously unavailable in the literature. Then we present the Half-normal 2TSF specification, where we assume that the two one-sided error components follow Half-normal distributions. An empirical study contrasts the two. In section III of the chapter we develop a Corrected OLS/Method of Moments (COLS/MM) estimator that allows us to estimate models where the density of the error term is not in closed form. In the context of this estimator we introduce unbiased estimators for higher-order central moments and cumulants of the error term in a regression (up to 5th), which we call the "kapa-statistics". In section IV of the chapter we present, the semi-Gamma 2TSF specification, where we assume that one of the two one-sided error components follows a Gamma distribution (while the other follows an Exponential). The density of this specification is not closed-form and so the application of the COLS/MM estimator is needed. We highlight this combination in the second empirical study of the chapter.

The chapter concludes with another new specification, the Generalized Exponential 2TSF one, where each one-sided error component is assumed to follow the distribution of the maximum of two i.i.d Exponentials. This specification was mainly developed in order to allow for non-zero modes of the marginal distributions of $w$ and $u$. Consequently we present measures and statistics that are based on the mode, rather than the usual expected value measures.

In chapter 4 we confront matters of statistical dependence: first, the case where the two one-sided components are dependent, a plausible assumption in many situations. We develop the 2TSF Correlated Exponential specification that allows for such dependence, as well as a formal statistical test for its existence. In section II of the chapter we deal with the case of regressor endogeneity, and we present in detail a Gaussian Copula estimation approach in order to account for it. As a how-to guide, this has wider applicability since it can be implemented in any regression setup, irrespective of the assumptions made on the error term of the regression.

Part C - Economic Applications includes chapters 5 and 6, where two applications of the 2TSF framework are presented. In chapter 5, we develop a Nash-bargaining model as a new contribution to the structural foundations that may lead to a 2TSF model. We then present an empirical application where we use the tools of chapter 4, namely the 2TSF Correlated Exponential specification with a Copula, to capture both the intra-dependence between the error components as well as regressor endogeneity.

In chapter 6, we take up the issue of management and we develop a 2TSF model of management in production, with the purpose of measuring its contribution in the output of a production process, treating it as a latent variable. This model has also the ability to explain and account naturally for the "wrong skewness" problem that is not-infrequently observed in production data samples. The chapter includes two empirical applications. The first implements the 2TSF Generalized Exponential specification and looks for correlations between the management contribution to output and the management score as measured according to the methodology developed by Nicholas Bloom and associates. The second showcases how inference can be conducted in samples that exhibit the "wrong skewness" issue.

Part D - References and Technical Appendices contains a unified list of references, which at times indicates the interdisciplinary flavor of this PhD thesis. It also contains the sometimes long and tedious Technical Appendices that include detailed derivations of the various results presented in the main text. We suspect that in some cases more efficient methods of arriving at the final results exist, but reaching the efficiency frontier is no easy matter.

In all things econometrics we deal exclusively with a cross-sectional setup, and we also maintain throughout the assumption of an identically and independently distributed sample.

The empirical applications that accompany the various chapters are more of an expository nature. Their main purpose is to show the modeling and estimation tools at work, rather than to fully explore the data samples.

The various mathematical results have been checked repeatedly, and in different time instances. Results were also confirmed by Monte Carlo simulation (like distribution
moments) and/or numerical calculation (for example, for all densities here we have verified numerically that they integrate to unity over their domain).

We use the terminology "density" instead of "probability density function" and "distribution function" instead of "cumulative distribution function". That's a $50 \%$ economy in words/characters.

Finally, regarding citations, we use the official year of the publication in which a scientific work appeared, and its place therein (volume, pages), when available. In today's on-line world many peer-reviewed works become officially available earlier, even at an earlier year. But since matters of precedence do not affect our work here, we use this "official" year date for cross-check purposes.--

## Chapter 1

## Literature

We review the papers that have used the two-tier stochastic frontier Model (2TSF) in applications or contributed theoretical results related to it. Since the literature on the subject is not large, we tried to make it as exhaustive as possible, as of July 2018. Portions of this review have been published in Papadopoulos (2015a).

## I. Applications of the 2TSF model.

The 2TSF model was introduced by Polachek \& Yoon (1987) with a focus on the labor market. The authors pointed to the fact that even in relatively homogeneous competitive labor markets we observe wage variation and not a single equilibrium wage as standard theory would predict.

Based on search theory premises, they attributed this phenomenon to the (optimal) existence of incomplete information: employees searching for work do not know of all the opportunities and work offers (because it would be costly or infeasible to obtain such information). On the other hand, employers do not know all workers that search for work, and what they would be willing to supply at any given wage. And because incomplete information is heterogeneous and varies from employer to employer and from employee to employee, we also observe wage dispersion.

In order to estimate both effects, the authors extended previous work by Hofler \& Polachek (1982) where a single-tier SF model with a negative one-sided term was used to estimate the effects of "employee's ignorance" only, to which they added an additional onesided unobservable term representing now the employer's ignorance and having a positive effect on the realized wage. And thus, the 2TSF model was born.

The authors assumed that the two one-sided error terms followed each an Exponential distribution, and derived the density of the three-component error term (the third one being the random disturbance, assumed to follow a zero-mean Normal distribution), under the assumption that the three components where jointly independent. By applying maximum
likelihood estimation the model, apart for the coefficients related to the wage determinants, provided also estimates for the parameters characterizing the distributions of the one-sided error-terms and therefore a quantitative assessment of their average effect on the realized wage. The authors also stratified their sample and calculated measures of employee and employer ignorance for various subsamples partitioned according to characteristics like gender, race, education and tenure. In all cases, the realized wage was on average below the "full-information" estimated level.

The stratification exercise in the foundational 2TSF paper accounted to a degree for the existence of heterogeneity that was used to rationalize the observed wage dispersion. 2TSF models that directly allowed for heterogeneity at the observation-level came later.

Groot \& Oosterbeek (1994), using an earnings-related data set from Netherlands, extended the model to be able to link informational inefficiencies to individual characteristics, by assuming that the moments of each one-sided disturbance is a function of the regressors and/or other variables. Since the distributions assumed have a single parameter, this setup not only allowed for individual heterogeneity regarding the mean value, but also accommodated conditional heteroskedasticity. As regards sample averages, the authors found that for their data sample the realized wages where above the full information wage.

Polachek \& Yoon (1996) extended their original model to accommodate panel data in order to disentangle unobserved individual heterogeneity from the informational imperfections. The authors applied a fixed-effects model and a two-step estimation procedure. They found that the panel data approach improved the quality of the results, and that the significance of incomplete information, although it was reduced, persisted.

Polachek (2017) digs more deeply into the relation between unobservable individual heterogeneity and incomplete information by combining the 2TSF model with research from Polachek, Das \& Thamma-Apiroam $(2013,2015)$ that exploited the fact that, as the years pass, long-enough time-series data on individuals are becoming available, permitting them to estimate five key individual parameters for a specific sample of individuals. Three of these parameters measure two types of ability, one quantifies skill depreciation, and the last one constitutes the respondent's time discount rate. Polachek (2017) estimates then a 2TSF model
for the same individuals, while using these five parameters to stratify the data and obtain how incomplete information may change in each subgroup.

A paper with an empirical application of the 2TSF model on the labor market is Sharif \& Dar (2007) which applied the original 2TSF model in cross-sectional earnings data from Canada focusing on immigrants. They found that realized wages were on average below the full-information level. In Dar (2014) the author used updated census data and extended the previous paper by using the heteroskedastic variant of the 2TSF model together with stratification.

Murphy \& Strobl (2008) applied the heteroskedastic 2TSF model to data from Trinidad \& Tobago. They found that actual wages hovered above the full-information wage. Methodologically, the authors conducted a specification test by applying also a Gamma specification (i.e. where the one-sided error components are assumed to follow Gamma distributions that nest the Exponential). This composite density has no closed-form and they used a simulated maximum likelihood approach. They reported that the Exponential specification was adequate.

Kumbhakar \& Parmeter (2009) proposed a different reason why variations around the full-information wage exist: they pointed out that the value of an employer-employee match is uncertain and remains so, and it is this uncertainty that creates the necessary space for bargaining to take place. And in a bargaining situation over some price, each side tries to pull the outcome in opposing directions: a framework suitable for the 2TSF approach ${ }^{1}$. Their model estimates the expected value of the match (the observable systematic component of the regression equation), and, through the two one-sided error components, the monetary value of the gap claimed by the two negotiating parties (depending on their relative bargaining power). Regarding quantitative findings, they obtained that the bargaining power of buyers (employers) was relatively higher than that of sellers (employees), leading the realized wage to be on average below the expected value of the match.

This model has seen applications to other markets and situations. Kinukawa \& Motohashi $(2010,2016)$ applied the model to the biotechnology market and the trading of biotechnologies/knowledge assets through company alliances/collaborations that are very

[^1]common in this market. They too found that buyers had greater bargaining power than sellers.

Wang Y (2016a) applied the bargaining model to the field of bilateral aid to developing countries. The literature on the topic has identified different setups in which aid takes place. One of them is the "aid-for-policy" framework, where aid is conditioned on the recipient countries implementing specific policies, either because the donors are self-interested and they require something in return, or because they are altruistic but have specific opinions as to what policies would benefit the population of the recipient country. Here the donors are the "buyers", "buying" the aid-recipients' government command over local resources and the authority to implement policies. The author found that donor countries enjoyed more bargaining power in surplus division than recipients. This conclusion was reinforced by the empirical study in Wang Y (2016b) related to USA economic aid for the period 1976-2011.

Zhang, Zhang, Yang \& Zhou (2017) applied the model to the Tourism industry in relation to tourist shopping. They found that tourists (buyers) extract a higher surplus compared to the sellers, which perhaps runs against widely held expectations that picture tourists as temporary customers, outside their comfort zone and their supporting social networks, and so in a "weaker" position than sellers. But on a second thought this has some intuition. Doing business with tourists is a high-volume short-length activity and so sellers are pressed to complete a high volume of transactions in a short period of time. This weakens their position in a bargaining situation, since they will bear the hidden cost of losing business if each individual negotiation is protracted. On the other hand tourists come in the negotiation with a bias that sellers will try to "rip them off" and so we may expect that they will put up a tough negotiating stance from the beginning.

These applications exemplify the wide reach of the 2TSF framework, outside the traditional topics of research in stochastic frontier analysis. But there is more.

The 2TSF model has also seen applications in the Health Services market, where informational asymmetries and inefficiencies are believed to abound, and to favor the supply side of the market. Gaynor \& Polachek (1994) applied it to the physician services market in USA arguing that the observed wide variations in physician fees went beyond differences in quality-of-service, and part of them were to be explained by incomplete and asymmetric information of the market participants. Their findings aligned with conventional wisdom,
estimating that the monetary effects of the incomplete information of patients were approximately $50 \%$ greater than those of the incomplete information of physicians. From a methodological point of view, the authors built their final regression specification by first formulating two separate equations with different regressors for the "reservation prices" of doctors and patients ("minimum accepted fee" and "maximum willingness to pay" respectively).

In a purely empirical paper Chawla (2002) used the original 2TSF model on health services data from Egypt, and found also that doctors were extracting a larger surplus in their transactions with the patients.

Tomini, Groot \& Pavlova (2012) estimated the effects of incomplete information on the informal ("under-the-table") payments made to physicians in Albania. Interestingly, they found that in this case the informational deficiencies were in favor of patients, against conventional wisdom which accepts that in the interaction between physicians and patients, the "balance of power" is in favor of the physicians. But the fact that, related to informal payments, the balance of power appears reversed has intuition and strengthens the case for the 2TSF model as a valid method to unearth existing phenomena: in the absence of a public market and price system, buyers-patients tend to exchange more information regarding informal costs, while on the other hand sellers-doctors possess less information for the same reasons, and also tend to avoid exchanging information because informal revenues are considered unethical and usually are illegal. The fundamental asymmetry here is the fact that patients are not saddled with the moral burden of participating in an illegal/unethical transaction -only physicians carry this burden.

Moving to other markets, Kumbhakar \& Parmeter (2010) developed a 2TSF hedonic price model for the house-selling market, and applied it to USA data. They implemented the standard model as well as the heteroskedastic variant. In both cases, on average, the selling price was somewhat below the full-information one, showing that buyers are in a better position than sellers. Rajapaksa (2015) applied their model to the housing market in Brisbane Australia, and on the contrary found that the incomplete information of buyers was higher than that of sellers, leading to a price above the full-information level. Tauer and Fried (2018) applied the model to study over-pricing and under-pricing in the wine market (of US Rieslings in the period 2000-2016).

Ferona \& Tsionas (2012) applied the model to an auction setting, using data from timber auctions, in order to assess the extent of systematic underbidding and overbidding behavior. They found that overbidding behavior tended to dominate, resulting in bids higher than optimal.

Researchers in China have taken a strong interest in the model, and applied it to diverse situations, with a strong focus in investment behavior${ }^{2}$ : Lian \& Chung (2008), Yu \& Liang (2012), Zhang \& Zheng (2012), Li, Wang and Zheng (2014) and Wen, Liu, Wang \& Caputo (2016) used it to investigate the effects of financing constraints and agency costs on investment behavior and also on dividend policies, of listed Chinese firms. Lv (2013) investigated volume and efficiency in R\&D investment of Chinese listed companies, pitting agency conflicts against incentive compatibility through a 2TSF model. The effects were reported as large but diminishing through time. Wei (2015) studied the opposing forces of financial constraints and government subsidies on R\&D investment. Lin, Liu \& Sun (2017) synthesized by examining the effects of financing constraints and agency costs on R\&D investment specifically. Liu (2017a, 2017b) examined the internal struggles in corporations that may lead to over-investment, while Xie and Li (2018) applied the model in order to measure investment efficiency and to test whether equity-incentives to management had an effect on the former.

Zheng \& Zhang (2012a), Huang (2013), Liu \& Liu (2014) used the model to separate the "premium" effect from the "under-pricing" effect in Initial Public Offerings (IPOs) of Chinese firms. All three studies found that the under-pricing effect dominated in the samples examined. Huang, Zia \& Wang (2017), went one step deeper and decomposed the underpricing effect into a discount effect from the primary market and a premium effect from the secondary market. Zheng \& Zhang (2012b) examined extreme IPO returns. Tao, Xin \& Lian (2014) examined the allocation of bargaining power between listed companies and banks in the credit market (and found that banks had the upper hand). Du and Wei (2014) used the model to measure the efficiency of urban industrial emissions. Zhang \& Sun (2015) investigated the exchange rate of China's currency RBM using the 2TSF model, and found evidence that, if anything, the intervention of the Chinese government in the exchange

[^2]market tends to overvalue the currency, contrary to the predominant belief. Xu, Wang, Zhou \& Geng (2016) identified a dual effect of government intervention on the real-estate market in China, one tending to increase prices and one tending to decrease them, and used the 2TSF model to quantify them. Yan and Qi (2017) used the 2TSF model to study the effects of asymmetric information and bargaining power in the fruit export market in China. Lyu, Decker \& Ni (2018) visit the labor market and examine the compensation of CEOs in Chinese firms using the 2TSF bargaining model of Kumbhakar and Parmeter (2009).

We have also found in the literature two exotic (and ingenious) applications of the 2TSF model.

Groot \& van den Brink (2007) used it to measure the effects of "optimism" and "pessimism" in self-reported quality of life. Compared to the estimated "realistic" (mean) values of life satisfaction, they found that "optimistic" people are too optimistic, while "pessimistic" people tend in comparison to be below of, but much closer to, the realistic value of life satisfaction. Other findings of their study were that men are relatively more optimistic and less pessimistic than women. Also, that cardiovascular disease makes people both less optimistic and less pessimistic, i.e. it dampens the intensity of these psychological tendencies, which is a reasonable result considering that people with such health issues are advised and usually do try to avoid strong emotional states.

Poggi (2010) went back to the labor market, but this time in order to measure "perceived job satisfaction" and how it is affected by downward and upward biases created by peoples' aspirations. She found that perceived job satisfaction languished on average a good $13 \%$ below its realistic level, a result that is consistent with the findings of the previous study: optimism (high aspirations) leads to disappointment and downward bias in evaluating the actual situation.

## II. Methodological advances related to the 2TSF model.

We turn now to look at papers that have contributed also at the technical/econometric level of the 2TSF framework, beyond the original distributional specification for cross-section and panel data and the heteroskedastic variant that have already been discussed.

The Kumbhakar \& Parmeter (2009) paper presented earlier, apart from introducing a 2TSF model for a bargaining situation, is also to be singled out at the technical/estimation level because it provided "refreshed" expressions for the likelihood of the composite error term, as well as formulas for individual metrics along the line proposed by Jondrow, Lovell, Materov \& Schmidt (1982). To a degree, it has succeeded the original Polachek \& Yoon (1987) paper as the new "application standard" for subsequent papers. ${ }^{3}$ Nevertheless, we must point out that the way the one-sided error components are obtained in their bargaining model, induces statistical dependence both between the two but also between the regressors and the composite error term, and so introduces endogeneity, something that is not taken into account in their econometric/statistical specification.

Only recently did papers appear that introduced new variants of the 2TSF model that allow us to handle different population structures and/or provide alternative estimation methods.

Papadopoulos (2015a) presented a new distributional specification for the composite error term, where the one-sided terms are assumed to follow each a Half-normal distribution instead of the Exponential one. ${ }^{4}$ The paper includes also formulas for individual, observation-specific measures for the one-sided error terms, either for a specification in levels or for the semi-log and log-log equation specifications that are commonly found in the literature. It also discusses statistical tests to validate the application of a 2TSF model, by exploiting either the implied skeweness of the composite error term, or its excess kurtosis in case it is close to symmetric (something that will hold if the two one-sided error terms tend to cancel each other out).

Blanco (2017) focused on the market for job placement services. He found that employees that used job placement services are not more informed about wage offers than employees that did not use the services, while firms that employed individuals that used placement services are more informed about reservation wages relative to firms that employed non-users (namely, job placement services tend to benefit more the employer side). On the methodological front, the author adjusted the 2TSF model to take into account

[^3]sample-selection bias, extending the simulated maximum likelihood methodology that Greene (2010) has developed for the single-tier SF framework.

In a purely methodological paper, Parmeter (2017) exploits the "scaling property" that characterizes all single-parameter distributions (and not only), in order to make feasible a non-linear least squares estimator (instead of the maximum likelihood one), thus doing away with the need for distributional assumptions that can cause misspecification. The approach has similarities with the heteroskedastic extension of the 2TSF model discussed previously, although here, the functional specification is non-linear to achieve identification in a leastsquares estimation framework. Another important property of this approach is that it allows automatically for the existence of dependence between the error terms and between them and the regressors, since in essence, the one-sided error terms become regressors themselves, leaving only the random symmetric disturbance as unobservable.

Das \& Polachek (2017a, 2017b) developed a new panel-data 2TSF framework in order to estimate gross flows in and out of the labor market, employment and unemployment ("Joiners and Leavers"), flows that are more important (compared to net ones) in order to understand the actual dynamics of the labor market. Although they assume that the onesided error components each follow an Exponential distribution as in the benchmark model, one of them is no longer identically distributed because it contains a group-specific heterogeneity parameter (so it remains an Exponential distribution but with changing parameter). This leads to a different likelihood function than the usual one, and a two-step estimation procedure. The model embeds heterogeneity directly into the composite error term of the 2TSF specification, and can be used in other settings. When panel-data are available, it offers a new way to disentangle unobserved heterogeneity from efficiency metrics.

Wang PY (2017) extends the approach of Kumbhakar, Lien \& Hardaker (2014) and formulates a six-component 2TSF model for panel data that includes an individual heterogeneity component, the random disturbance, and where each one-sided error term is decomposed into a time-varying and a time-invariant part. This model nests most of the single-tier and two-tier SF models in the literature. The author presents also the Truncated

Normal 2TSF cross-sectional specification, where the one-sided terms follow the Truncated Normal distribution, but unfortunately the provided expression for the density is wrong. ${ }^{5}$

Finally, in what may be an important new direction, Huang, Luo \& Wang (2018) combine the 2TSF approach with a business cycle model with autocorrelation, opening at once two new territories for the model, macroeconomics and the handling of serial correlation.

## III. "Double frontiers" in Data Envelopment Analysis. ${ }^{6}$

We close this literature review by looking at approaches that have been presented in the cousin field of Data Envelopment Analysis (DEA) and are similar in spirit to the 2STF framework.

Lins, de Lyra Novaes \& Legey (2005) introduced the "Double Perspective" DEA model (DP-DEA), in order to study the housing market in Brazil but also to provide a real-estate value assessment tool. In an ingenious move, the authors essentially created the equivalent of the core of an Edgeworth's exchange box by taking an input-oriented DEA model, transposing it and super-imposing it on an output-oriented DEA model. Here, we do not look at pair of goods as in the original Edgeworth box, but at combinations of value-area of real-estate properties that, given the data, appear to be the "feasible" combinations (in an economic and not technical sense). With price on the vertical axis, the "concave-down" border of this core is the "seller's efficiency frontier", while the "concave-up" border of the core is the "buyer's efficiency frontier", in the sense of maximum surplus extraction.

Hadley and Ruggiero (2006) applied the same approach to study the arbitraged salary negotiations in the market for Major League Baseball players in USA.

Mouchart and Vandresse $(2007,2010)$, studied the freight market in Belgium, modeling the contract space of the related negotiations as a "maximum willingness to pay/minimum willingness to sell" double frontier, an approach equivalent to that of Gaynor and Polachek (1994) (see chapter 2). But here, the estimation method is an extension of the standard DEA approach using "biderictional" free disposability. It is interesting that the estimated densities (p. 1301, figure 3 in the 2007 paper) of what in their model corresponds to the positive and

[^4]negative one-sided error components in the 2TSF approach roughly indicate an Exponentiallike distribution for both.

Lakhdar, Leleu, Vaillant \& Wolff (2013) study the illicit drug trade. They abandon the convexity assumption of DEA and formulate a double-frontier model using a Free Disposal Hull (FDH) model. They explicitly invoke incomplete information as the force that drives price up or down from a perfect-information equilibrium, and estimate the distribution of the inefficiency components for four different drugs (herbal cannabis, cannabis resin, crack cocaine, powder cocaine). Here the estimated distributions for both sellers and buyers vary considerably from drug to drug. At a second estimation stage, they regress the series of buyer's inefficiency scores on personal attributes like gender, age and occupation, as well as on indicators of experience with the drug (years involved in using it, daily quantity consumed). This is similar in spirit to the 2TSF Scaling Property approach proposed by Parmeter (2017), where co-variates are used as determinants of inefficiencies.

Wolff (2016) adopts also an FDH approach to study "bargaining power" in on-line diamond markets. The author first runs a standard hedonic-price regression in order to select the most important features that influence the buying-selling decision, which are subsequently used in his main model.

Yet in another direction extending DEA, Shabanpur, Yousefi \& Saen (2017) apply a multistage Goal Programming-DEA model to estimate a double-frontier, this time around the same "decision making unit" (this is structurally analogous to the situations examined by Groot \& van den Brink 2007 and Poggi 2010). Their case study concerns estimating the "sustainability" of a sample of suppliers, creating in the process upper and lower boundaries for (in)efficiency using data on inputs like price, environmental and work-safety indices, and on outputs like quality of products, financial stability and efficiency in energy consumption.

It is natural to envision a comparative study of these methods together with the 2TSF approach, in the sense of applying all of them on the same data set(s) and explore and understand the differences in the obtained results (or marvel at their similarity). This would best be a collaborative effort, and it is indeed a direction of research that we plan to take in the future.--

## Chapter 2

## Structural Foundations

We analyze in detail the different structural foundations for the 2TSF model that have been put forth. We identify certain conceptual and statistical issues that arise, and we propose refinements to improve their soundness and interpretative power. We find one of them to have a critical weakness that makes it an invalid base for a 2TSF model. The chapter closes with a general "latent variables" argument for the use of the 2TSF approach.

## I. The "incomplete information" framework of Polachek \& Yoon

## (1987).

In the first paper to introduce the 2TSF model, Polachek \& Yoon (1987) (P\&Y-1987 thereafter) started their modeling of the labor market by treating labor demand and supply schedules as representing maximum quantities demanded and supplied ${ }^{1}$. Namely, they are frontiers. As such the actual observed demand and supply functions were stochastically modeled as follows:

$$
\begin{align*}
& L^{D}=f\left(\mathbf{x}^{D}, \omega\right)-e^{D}, \quad \partial f / \partial \omega<0,  \tag{2.1}\\
& L^{S}=g\left(\mathbf{x}^{S}, \omega\right)-e^{S}, \quad \partial g / \partial \omega>0, \tag{2.2}
\end{align*}
$$

where $\omega$ equals wage; $e^{D}$ is a non-negative random variable, reflecting the fact that the actual quantity demanded, $L^{D}$ is below $f\left(\mathbf{x}^{D}, \omega\right)$, the maximum quantity demanded at wage $\omega$; the vector $\mathbf{x}^{D}$ lists the determinants of $f\left(\mathbf{x}^{D}, \omega\right)$ other than $\omega$. Similarly for the supply curve, the actual quantity supplied $L^{S}$ is below the maximum labor quantity

[^5]$g\left(\mathbf{x}^{s}, \omega\right)$ which will be supplied at any wage level, $\omega$. The term $e^{s}$ is also a non-negative random variable. In what sense "actual quantity demanded or supplied is below its corresponding maximum" is a crucial conceptual point that we will investigate shortly.

The authors then defined the non-stochastic portion of excess labor demand,

$$
\begin{equation*}
h(\mathbf{x}, \omega) \equiv f\left(\mathbf{x}^{D}, \omega\right)-g\left(\mathbf{x}^{S}, \omega\right), \quad \mathbf{x}=\left(\mathbf{x}^{D}, \mathbf{x}^{S}\right) \tag{2.3}
\end{equation*}
$$

and imposed the equilibrium market-clearing condition $L^{D}=L^{S}$, under which we obtain

$$
\begin{equation*}
h\left(\mathbf{x}, \omega^{*}\right)=e^{D}-e^{s}, \tag{2.4}
\end{equation*}
$$

where the star denotes equilibrium wage. Subsequently, they applied a first-order Taylor expansion on $h\left(\mathbf{x}, \omega^{*}\right)$ around a fixed reference point $\left(\mathbf{x}_{0}, \omega_{0}\right)$ at which $h\left(\mathbf{x}_{0}, \omega_{0}\right)=0$, i.e. around the point where the labor market clears in the absence of the stochastic terms $e^{D}, e^{S}$ (or when it so happens that $e^{D}=e^{S}$ ). But this means that the vector $\left(\mathbf{x}_{0}, \omega_{0}\right)$ represents an equilibrium at the full-information wage, $\left(\mathbf{x}_{0}, \omega_{0}\right)=\left(\mathbf{x}_{F I}, \omega_{F I}\right)$, although most likely not at full-information labor employed, since $h\left(\mathbf{x}_{F I}, \omega_{F I}\right)=0$ requires only that $e^{D}=e^{S}$ and not that $e^{D}=e^{S}=0$. The Taylor expansion reads

$$
h\left(\mathbf{x}, \omega^{*}\right)=h\left(\mathbf{x}_{F I}, \omega_{F I}\right)+\left(\mathbf{x}-\mathbf{x}_{F I}\right)^{\prime} \nabla_{\mathrm{x}} h\left(\mathbf{x}_{F I}, \omega_{F I}\right)+\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\left(\omega^{*}-\omega_{F I}\right)+R_{h} .
$$

Using $h\left(\mathbf{x}_{F I}, \omega_{F I}\right)=0$ and eq. [2.4] we obtain

$$
\begin{equation*}
e^{D}-e^{S}=\left(\mathbf{x}-\mathbf{x}_{F I}\right)^{\prime} \nabla_{\mathrm{x}} h\left(\mathbf{x}_{F I}, \omega_{F I}\right)+\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\left(\omega^{*}-\omega_{F I}\right)+R_{h} . \tag{2.5}
\end{equation*}
$$

Note that

$$
\frac{\partial h(\mathbf{x}, \omega)}{\partial \omega}=\frac{\partial f\left(\mathbf{x}^{D}, \omega\right)}{\partial \omega}-\frac{\partial g\left(\mathbf{x}^{S}, \omega\right)}{\partial \omega}<0
$$

due to the signs of the two derivative components. Attaching a minus sign to this derivative and re-arranging to obtain an expression for $\omega^{*}$ we get

$$
\omega^{*}=\left(-\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\right)^{-1}\left[\begin{array}{r}
-\mathbf{x}_{F I}^{\prime} \cdot \nabla_{\mathrm{x}} h\left(\mathbf{x}_{F I}, \omega_{F I}\right)+\frac{-\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega} \omega_{F I}  \tag{2.6}\\
+\mathbf{x}^{\prime} \cdot \nabla_{\mathrm{x}} h\left(\mathbf{x}_{F I}, \omega_{F I}\right)+R_{h}+e^{s}-e^{D}
\end{array}\right] .
$$

This provides the mapping to the components of an econometric regression specification,

$$
\begin{gather*}
\omega^{*}=\beta_{0}+\mathbf{x}^{\prime} \boldsymbol{\beta}+v+w-u  \tag{2.7}\\
\beta_{0}=\omega_{F I}-\mathbf{x}_{F I}^{\prime} \cdot \nabla_{\mathrm{x}} h\left(\mathbf{x}_{F I}, \omega_{F I}\right) \cdot\left|\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\right|^{-1}, \quad \boldsymbol{\beta}=\left|\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\right|^{-1} \nabla_{\mathrm{x}} h\left(\mathbf{x}_{F I}, \omega_{F I}\right) . \tag{2.8}
\end{gather*}
$$

The equation includes three distinct "error" components: $v$ that maps to the scaled Taylor remainder $\left(\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right) / \partial \omega\right)^{-1} R_{h}$ and to other purely random components we should allow for, while the two components that interest us are

$$
\begin{equation*}
w=\left|\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\right|^{-1} e^{s}, \quad u=\left|\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\right|^{-1} e^{D} . \tag{2.9}
\end{equation*}
$$

Since both $e^{S}, e^{D}$ are non-negative random variables, and the second enters with a negative sign, the specification [2.7] has the error structure of the 2TSF model.

## I.1. Interpretation of the supply and demand stochastic shifters.

From eq. [2.9] we see that a higher $e^{s}$, which from eq. [2.2] tends to decrease (shift upwards) the labor supply, is associated with higher equilibrium wage, while a higher $e^{D}$, which from eq. [2.1] tends to decrease (shift downwards) labor demand, is associated with a lower equilibrium wage.

The question to be asked now is: what these two components may represent? P\&Y1987, using a Search theory approach, mapped them to (optimal due to information acquisition costs) incomplete information in the labor market (its pecuniary effects on the wage, to be precise). So, they argued, $u$ represents the effects of "worker ignorance", showing how much less workers receive than the full-information wage, while $w$ represents how much more workers receive than the full-information wage, and it can be thought of as the consequence of "employer's ignorance".

But looking at eq. [2.9] we see that $u$ (the "worker's ignorance") includes the $e^{D}$ term, which represents the downward influence on the labor demand function which is the employer side of the market. It is a term that originally exerts quantity effects on demand, and not price effects (it is transformed into price/value in the equilibrium wage equation by the derivative of the excess demand function). Analogously the $e^{S}$ term is present in the supply curve (worker's side) but represents employer's effects of incomplete information.

So, in order to conceptually validate this structural derivation of the 2TSF model at the market level we have to ask and answer the following questions: In what sense worker's incomplete information reduces quantity of labor demanded by employers? In what sense employer's incomplete information reduces quantity of labor supplied by the workers?

Since the terms $e^{S}, e^{D}$ are defined as quantity shifters, to rationalize the relations we think of incomplete information as "not knowing the full extent of the market", without inserting reservation wages into the picture: workers do not know all individual quantities demanded by each and every firm at any given wage level, and so neither do they know their sum, and firms do not know all individual quantities supplied at any given wage level, neither their sum.

This means that the labor supply curve that is present in the market and interacts with demand is effectively lower (upward shift) than the labor supply schedule that represents the sum of the quantities supplied by all workers. And this happens due to incomplete
information of employers. So the $e^{s}$ term is a quantification of the labor supply that stays "invisible", but does so due to employers' ignorance.

Likewise, the labor demand schedule that is "active" in the market and interacts with supply is effectively lower (downward shift) than the labor demand schedule that represents the sum of the demands of all firms. And this happens due to incomplete information of workers. So the $e^{D}$ term is the quantification of the labor demand that stays "invisible", but does so due to "worker's ignorance".

Wording the situation as an aphorism, we could say "If the other side does not know that I exist, I don't". Diagrammatically we have

Figure 1: Incomplete information at market level as failure to know the whole market.


The model is consistent with incentives to reduce the degree of incomplete information: for employers, "seeing" a larger portion of supply (smaller $e^{S}$ ) means lower
wage, higher employment and higher output. For workers, "seeing" a larger portion of labor demand (smaller $e^{D}$ ) means a higher wage (and higher quantity employed) at equilibrium.

From the diagram it is clear that the unambiguous effect of incomplete information is to depress the equilibrium employment level. The effect on the equilibrium wage is ambiguous: it may be higher, lower or equal to the full-information wage. This means that at the observed wages we will have two price frontiers, as we have described in abstract in the Introduction of this thesis.

Moreover, if the actual wage is above the full-information level, then the workers' side will experience excess labor supply: at that wage more workers are willing to sell their labor, but they do not find employers. If the actual wage is below the full information wage, the employers' side will experience excess labor demand: they would be willing to hire more at this wage, but they do not find workers.

## I.2. Wage dispersion.

The goal of P\&Y-1987 was to use incomplete information in order to explain wage dispersion even in relatively homogeneous labor markets, rather than just depressed employment, as the above single-equilibrium wage diagram depicts. But this is not a big step to take, if we realistically assume that otherwise similar firms and similar workers are heterogeneous regarding their incomplete information, which we remind exists due to search costs. This informational heterogeneity fractures the market since not all firms meet with all workers/prospective employees, in effect creating many diagrams as the above, each being characterized by possibly different $e^{S}, e^{D}$ terms, and so each reaching a possibly different equilibrium/realized wage, in the context of the "same" labor market.

The incomplete information interpretation has received empirical validation in Polachek \& Robst (1998). The authors used independent direct measures of workers' "knowledge of the world of work" obtained from the National Longitudinal Survey of Young Men (NLSYM) in USA. They compared frontier estimates of incomplete information to these direct measures of workers' knowledge and verified that stochastic frontier analysis provides a reasonable measure of a worker's incomplete information.

## I.3. Interpretation of the regression equation and its coefficients.

We return to our regression specification. Looking at expression [2.8] that contains the mapping of the regression coefficients to the underlying structural components we can deduce the following: consider the total differential

$$
d h(\mathbf{x}, \omega)=\left[\nabla_{\mathrm{x}} h(\mathbf{x}, \omega)\right]^{\prime} d \mathbf{x}+\frac{\partial h(\mathbf{x}, \omega)}{\partial \omega} d \omega
$$

and set it equal to zero,

$$
d h(\mathbf{x}, \omega)=0 \Rightarrow d \omega=\left|\frac{\partial h(\mathbf{x}, \omega)}{\partial \omega}\right|^{-1} \nabla_{\mathrm{x}} h(\mathbf{x}, \omega)^{\prime} d \mathbf{x}, \quad \forall(\mathbf{x}, \omega) .
$$

So we also have

$$
\begin{equation*}
d h\left(\mathbf{x}_{F I}, \omega_{F I}\right)=0 \Rightarrow d \omega_{F I}=\left|\frac{\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right)}{\partial \omega}\right|^{-1} \nabla_{\mathrm{x}} h\left(\mathbf{x}_{F I}, \omega_{F I}\right)^{\prime} d \mathbf{x}_{F I} . \tag{2.10}
\end{equation*}
$$

Combining with [2.8] we see that

$$
\begin{equation*}
d \omega_{F I}=\left.\boldsymbol{\beta}^{\prime} d \mathbf{x}_{F I}\right|_{d h\left(\mathbf{x}_{F I}, \omega_{F I}\right)=0} . \tag{2.11}
\end{equation*}
$$

Eq. [2.11] illuminates the interpretation of the regression slope coefficients. They represent marginal effects of the regressors on the dependent variable, the wage, so that fullinformation excess labor demand remains zero, i.e. so that the market clears absent the incomplete information effects (or when they offset each other in a way to maintain fullinformation wage). Therefore this will be exact if we are at an equilibirum where the wage is at its full-information level, although labor employed may not be.

Moreover, each beta is not just the partial effect on the dependent variable (i.e. valid only if we set all but one of the differentials of the regressors equal to zero), it is the total marginal effect: the "required response" of the wage remains the same per regressor being changed, even if we allow more than one regressor to change at the same time, as long as we start pivoting from a full-information wage-point.

## The model at the observation level.

All the analysis of the model up to now has used market-level magnitudes since the regression specification was derived starting from market supply and demand. To transfer it to the individual level, we first re-write for individual transaction $i$ the econometric specification [2.7] using [2.8] in a more intuitive way,

$$
\begin{equation*}
\omega_{i}^{*}=\omega_{F I}+\left(\mathbf{x}_{i}-\mathbf{x}_{F I}^{R}\right)^{\prime} \boldsymbol{\beta}+v_{i}+w_{i}-u_{i} . \tag{2.12}
\end{equation*}
$$

The symbols must now be re-interpreted. The left-hand side, $\omega_{i}^{*}$, is no longer the "market equilibrium wage", not even the equilibrium wage of some sub-segment of the market. It is an individual wage. In the right-hand side we have the full-information market wage $\omega_{F I}$, and the term $\left(\mathbf{x}_{i}-\mathbf{x}_{F I}^{R}\right)^{\prime} \boldsymbol{\beta}: \mathbf{x}_{F I}^{R}$ is the regressor vector compatible with full-information wage for the "representative" worker and firm. The vector $\mathbf{x}_{i}$ represents the actual characteristics of the firm/worker pair of observation $i$. So the systematic part of the regression reflects how an individual wage deviates from the full-information market wage $\omega_{F I}$ due to individual heterogeneity and dispersion of characteristics. We also have to re-interpret the one-sided error components: now they reflect incomplete information about the other party of observation $i$, as it may relate to minimum wage acceptable, maximum wage payable, and other characteristics. Given these, averaging over $i$ we obtain

$$
E\left(\omega^{*}\right)=\omega_{F I}+\left(E\left(\mathbf{x}_{i}\right)-\mathbf{x}_{F I}^{R}\right)^{\prime} \boldsymbol{\beta}+E\left(w_{i}\right)-E\left(u_{i}\right) .
$$

But an intuitive understanding of "representative" is the mean/expected value, so that $E\left(\mathbf{x}_{i}\right)=\mathbf{x}_{F I}^{R}$, and therefore we are left with

$$
\begin{equation*}
E\left(\omega^{*}\right)=\omega_{F I}+E\left(w_{i}\right)-E\left(u_{i}\right) . \tag{2.13}
\end{equation*}
$$

The sample analogue is

$$
\begin{equation*}
\frac{1}{n} \sum_{i=1}^{n} \omega_{i}^{*}=\hat{\omega}_{F I}+\hat{E}(w)-\hat{E}(u), \tag{2.14}
\end{equation*}
$$

where $\left\{\omega_{i}^{*}\right\}$ is the observed series of the dependent variable. Applying maximum likelihood estimation we can obtain the estimates $\hat{E}(w), \hat{E}(u)$ and so arrive also at an estimate for the full information wage

$$
\begin{equation*}
\hat{\omega}_{F I}=\frac{1}{n} \sum_{i=1}^{n} \omega_{i}^{*}-[\hat{E}(w)-\hat{E}(u)] . \tag{2.15}
\end{equation*}
$$

Note that under the relation $E\left(\mathbf{x}_{i}\right)=\mathbf{x}_{F I}^{R}$, we can empirically implement directly eq. [2.12] by centering the regressors on their mean value (but not the dependent variable).

## I.4. Statistical and econometric issues. ${ }^{2}$

The model includes in its "random" disturbance $v$ the scaled Taylor remainder $\left(\partial h\left(\mathbf{x}_{F I}, \omega_{F I}\right) / \partial \omega\right)^{-1} R_{h}$ from the linearization of the equilibrium non-stochastic portion of excess labor demand $h\left(\mathbf{x}, \omega^{*}\right)$. At the market level this represents the discrepancy due to the fact that, at the actual regressor vector $\mathbf{x}$, the corresponding wage that makes excess demand equal to zero (even if the incoplete information terms are absent) may be different, and so linked

[^6]to different equilibrium-restoring marginal effects. At individual level, it represents the discrepancy of the actual firm-worker regressor vector from the representative one, $\mathbf{x}_{F I}^{R}$.

As such the Taylor remainder will be a function of all other regressors, indicating the possible existence of regressor endogeneity.

In fact, endogeneity is the price to pay for abandoning OLS estimation. One can show that, if we treat the systematic linear part of the regression as a first-order approximation to the possibly non-linear conditional expectation function of the wage given the regressors, then, if a) we consider the Taylor expansion as taken around the expected values of the regressors and b) we estimate by Ordinary Least Squares, then, the Taylor remainder becomes uncorrelated to the regressors (at the limit), and so the estimator is still consistent.

But in our case only the first condition is met. We showed that we can realistically view the Taylor expansion as being taken around the expected value of the regressors, but we do not estimate the model by OLS, but by maximum likelihood on a not-normal error density.

So while OLS remains consistent (for the slope coefficients although not for the constant term), the MLE will likely not be, contaminated by regressor endogeneity. But we are relying on the MLE to obtain estimates for the constant term as well as for the parameters of the one-sided error components, and we went to all this trouble exactly for these three pieces of information, namely the full-information wage (the constant term) and the deviations from it due to the incomplete information (the two one-sided error components).

Another issue regarding the econometric implementation of the model is that it describes a regressor matrix that includes characteristics of both sides of the market, employers and workers, as well as of the market itself. This would require a matched employer-employee data set, which in many cases is unavailable, and what we have are regressors pertaining to the worker-side only, plus perhaps some market characteristics like for example geographical or socio-economic characteristics of the environment (urban, country, industry, etc). So in most circumstances, the information contained in the regressor matrix will be unbalanced, leaving in the "error term" a host of regressors related to the employer-side, which, if correlated with the included regressors, will intensify the endogeneity problem. So we see that the original formulation of the two-tier stochastic frontier model is most likely characterized by regressor endogeneity and it should be estimated by a method that takes into account this dependence structure.

And even if this is not the case, we still have omitted variables in the model, meaning that the estimate of the constant term will lose its interpretation as the full-information wage, since it will incorporate also the means of these omitted variables that represent employercharacteristics.

## II. The "reservation price" framework of Gaynor \& Polachek (1994).

This was the first paper to apply the 2TSF model to the Health Services market, a natural choice due to the informational asymmetries and imperfections that characterize it.

The authors (G\&P-1994 thereafter) start their build-up towards the 2TSF reduced-form by assuming structural regression equations for the reservation prices of both buyer (patient) and seller (physician). Namely, here the model starts at the individual level, imagining a bilateral transaction between buyer and seller.

The reservation-price equation for the buyer is

$$
\begin{equation*}
P_{b}=\mathbf{x}_{b}^{\prime} \beta_{b}+v_{b}, \tag{2.16}
\end{equation*}
$$

where $\mathbf{x}_{b}^{\prime}$ contains the factors that affect the maximum fee the patient is willing to pay (such as the extent of insurance coverage, the patient's education and income, the severity of the patient's illness, and the frequency with which the physician's services are needed), and $v_{b}$ is a random disturbance. Analogously for the physician, we have

$$
\begin{equation*}
P_{s}=\mathbf{x}_{s}^{\prime} \beta_{s}+v_{s} \tag{2.17}
\end{equation*}
$$

where $\mathbf{x}_{s}^{\prime}$ contains regressors such as input prices, technology, age of equipment, and factors affecting efficiency.

G\&P-1994 then define "gains" for the buyer and the seller, as the distance between reservation prices and actual price paid $P_{c}$, or price received $P_{r}$. Such distances are a consequence of incomplete information in both sides of the transaction. The gain for the seller is defined as

$$
\begin{equation*}
W \equiv P_{c}-P_{s} \geq 0 \tag{2.18}
\end{equation*}
$$

i.e. the amount the buyer/patient pays above the reservation price of the physician (so this is due to "buyer's ignorance"). The gain for the buyer is defined as

$$
\begin{equation*}
U \equiv P_{b}-P_{r} \geq 0 \tag{2.19}
\end{equation*}
$$

i.e. the difference of the amount the buyer/patient pays from the maximum price he would be willing to pay (so this is due to "supplier's ignorance").

For a transaction to occur, price paid has to be equal to price received $P_{c}=P_{r}=P$. using this and combining equations [2.16] with [2.19] and [2.17] with [2.18] we get

$$
\begin{align*}
& U=\mathbf{x}_{b}^{\prime} \beta_{b}+v_{b}-P \Rightarrow P=\mathbf{x}_{b}^{\prime} \beta_{b}+v_{b}-U  \tag{2.20}\\
& W=P-\mathbf{x}_{s}^{\prime} \beta_{s}-v_{s} \Rightarrow P=\mathbf{x}_{s}^{\prime} \beta_{s}+v_{s}+W \tag{2.21}
\end{align*}
$$

We have obtained two different expressions for the observed transaction price. Each could be used as a single-tier stochastic frontier model, eq. [2.20] to measure the gain to the buyer due to seller's ignorance, while eq. [2.21] to measure the gain to the seller due to buyer's ignorance (note the different set of regressors). In order to combine them into a 2TSF formulation, P\&G-1994 summed up both equations and divided by 2 , resulting in

$$
\begin{equation*}
P=\mathbf{x}^{\prime} \beta+v+w-u \tag{2.22}
\end{equation*}
$$

where

$$
\mathbf{x}^{\prime}=\left(\mathbf{x}_{s}^{\prime}, \mathbf{x}_{b}^{\prime}\right), \beta=\frac{1}{2}\left(\beta_{s}, \beta_{b}\right)^{\prime}, \quad v=\frac{1}{2}\left(v_{s}+v_{b}\right), \quad w=W / 2, u=U / 2 .
$$

Eq. [2.22] has the structure of a 2TSF reduced-form equation.

It is important to point out that the definitions for $W$ and $U$ (eq. [2.18] and [2.19] respectively) are ex post representations, and not ex ante structural relations. If they were the latter, they would render eq. [2.22] an identity, let alone inducing all sorts of statistical dependencies that would make the distributional specification seriously misspecified. In other words, $W$ and $U$ are only measured ex post as indicated by the right-hand sides of eq. [2.18] and [2.19], they are not caused by these expressions, but rather, they arise due to the incomplete information of the participants in the transaction.

This is a foundation for the 2TSF model that can be used for any market/transaction where we can argue for the existence of reservation prices on both sides of the market. Here too the richness of the regressor matrix will determine how close the obtained estimates will be to their theoretical selves (in a conceptual way).

Regarding interpretation of the estimation results using [2.22] as obtained, in general the structural parameters will not be identifiable but only their average or their net value. This will depend on whether a regressor is present in both the underlying equations, or in only one of them. Finally, we should not forget the factor $1 / 2$ when we quantitatively assess the effects of incomplete information on price, since in expected-value terms we have $\hat{E}(W)=2 \hat{E}(w), \hat{E}(U)=2 \hat{E}(u)$, and it is $\hat{E}(w), \hat{E}(u)$ that we will obtain from the estimation procedure, while it is $E(W), E(U)$ that we are interested in.

## III. The "hedonic price" framework of Kumbhakar \& Parmeter

(2010).

In this paper the authors (K\&P-2010 thereafter) applied the 2TSF model in the houseselling market, in a hedonic analysis framework. Superficially, their approach may appear essentially similar with the one in Gaynor \& Polachek (1994) analyzed just above, but it is not, quite the contrary, and it leads to different quantitative consequences.

K\&P-2010 define the gains to the buyer and the seller due to the incomplete iformation of the other in the same way as in Gaynor \& Polachek (1994), but they do not construct structural equations for the reservation prices of buyers ("willingness to pay") and sellers ("willingness to accept").

When they impose the necessary condition that price received must equal price paid, they essentially add the loss to the seller and subtract the gain to the buyer to the transaction price, thus creating a "full-information" price expression. In the notation of the previous part we have

$$
\begin{equation*}
P_{F I}=P+u-w, \quad u=U, w=W . \tag{2.23}
\end{equation*}
$$

They then point out that in a hedonic analysis approach we have the hedonic function decomposition of full-information price (not actual price)

$$
\begin{equation*}
P_{F I}=h(\mathbf{z})+v, \tag{2.24}
\end{equation*}
$$

where $\mathbf{z}$ is a vector of characteristics of the house on sale, $h(\cdot)$ is the hedonic function, and $v$ is a random disturbance. Equating the two and rearranging we get

$$
\begin{equation*}
P=h(\mathbf{z})+v+w-u, \quad u=U, w=W, \tag{2.25}
\end{equation*}
$$

which is a 2TSF reduced form equation.
Obviously, the critical difference from the "reservation price" framework of Gaynor \& Polachek (1994) analyzed previously is that here the one-sided error terms in the 2TSF reduced form equal the gains of the parties due to incomplete information, while previously each was only half of these magnitudes (compare eq. [2.25] with eq. [2.22]). This is certainly crucial when using the model to obtain quantitative results. So it is important to understand clearly why do these frameworks differ, and so when it is appropriate to use the one or the other.

The fundamental difference is that K\&P-2010 obtain two equations for the fullinformation price, something that allows them to eliminate it and obtain a single expression for the transaction price. Gaynor \& Polachek (1994) do not assume the existence of an "independent" expression for the "full information" price (physician's fee) as do K\&P-2010 (the hedonic equation). But they do assume the existence of structural equations for the reservation prices (while K\&P-2010 do not). Consequently, what G\&P-1994 obtain is two
expressions for the transaction price (eq. [2.20] and eq. [2.21]) and they have to add them and divide by 2 to arrive at a single expression that can be implemented econometrically.

Both frameworks are valid. As for which one to use, it will depend on the data available and the model developed in each case. The first full-information price expression, eq. [2.23] can be directly assumed in any case by just arguing that incomplete information exists. If the researcher also has a model where an additional full-information price equation can be formed in terms of other variates, he should then apply the Kumbhakar \& Parmeter (2010) approach, and use the mapping $w=W, u=U$. If no such second full-information equation is available, and instead the researcher can build equations for the reservation prices, he should apply the Gaynor \& Polachek (1994) framework and use the mapping $w=W / 2, u=U / 2$.

## IV. The "Nash bargaining" framework of Kumbhakar \& Parmeter

 (2009).In this paper the authors built a 2TSF model by starting from the equilibrium wage equation obtained in the context of the benchmark Search \& Match labor market model of Pissarides (2000, ch. 1) that uses a Nash bargaining solution concept.

It is useful to start with Pissarides' equilibrium wage equation (his eq.1.18), using the author's notation: $w=r U+\beta(p-r U)$, where $w$ is the wage, $r U$ is the worker's reservation wage (related to unemployment benefits), $p$ is gross output obtained from the job filled by the worker, and $0 \leq \beta \leq 1$ is the relative bargaining power of the worker. It is important to remember that this model is non-stochastic. K\&P-2009 re-wrote the above equation as follows:

$$
\begin{equation*}
\omega=\underline{\omega}+\eta(\bar{\omega}-\underline{\omega}), \tag{2.26}
\end{equation*}
$$

where $\omega$ is the equilibrium wage, $\underline{\omega}$ is the reservation wage of the worker, $\eta=\beta$ and $\bar{\omega}$ is the firm's maximum wage offer. So K\&P-2009 mapped the certain gross output $p$ of

Pissarides to the maximum amount that the firm is willing to pay as wage to the employee. In a deterministic context this is indeed the reasonable correspondence -the most a firm would be willing to pay is what will be produced by the worker filling the position.

But the authors were interested in introducing productivity uncertainty that moreover persists even at the end of negotiations and the consummation of the contract ${ }^{3}$. So they defined the expected output $p$ of the match conditional on a vector of observed characteristics which, as the authors write, "are certainly used in hiring decisions by firms", $E(p \mid \mathbf{x})=\mu(\mathbf{x})$. Then, by using an add-and-subtract technique transformed eq. [2.26] to a 2TSF format,

$$
\begin{equation*}
\omega=\mu(\mathbf{x})+\eta(\bar{\omega}-\mu(\mathbf{x}))-(1-\eta)(\mu(\mathbf{x})-\underline{\omega}) . \tag{2.27}
\end{equation*}
$$

Just prior to deriving eq. [2.27], the authors wrote (quote) "by construction $\underline{\omega} \leq \mu(\mathbf{x}) \leq \bar{\omega}$ for those matches where a job is consummated". But in a stochastic environment where uncertainty is not resolved prior to consummate the hire, the inequality $\mu(\mathbf{x}) \leq \bar{\omega}$ appears conceptually problematic.

Faced with uncertainty as to what the productivity of the match will be, the firm must form a prediction about it , and the conditional expectation based on available information, while not the only one, is the best predictor in mean-squared error sense, and widely used as predictor. So during the negotiations, if the firm is asked "what is the gross output that you expect from this match", it will likely answer " $\mu(\mathbf{x})$ " (as defined above). And given this answer, the rational thing to ascertain is that the firm's maximum wage offer will be $\mu(\mathbf{x})$, $\bar{\omega}=\mu(\mathbf{x})$, and certainly never $\bar{\omega}>\mu(\mathbf{x})$. So, $\bar{\omega}=\mu(\mathbf{x})$ appears to be the correct analog in a stochastic environment, to equating $\bar{\omega}$ to Pissarides' gross output $p$ in a certain environment. But then, with $\bar{\omega}=\mu(\mathbf{x})$ eq. [2.27] collapses to

$$
\begin{equation*}
\omega=\mu(\mathbf{x})-(1-\eta)(\mu(\mathbf{x})-\underline{\omega}) \tag{2.28}
\end{equation*}
$$

[^7]The term $\mu(\mathbf{x})-\underline{\omega}$ is the total distance over which to bargain, given persistent productivity uncertainty. Weighted by the firm's relative bargaining power $1-\eta$, and being subtracted from $\mu(\mathbf{x})=\bar{\omega}$, it expresses how much the firm gains due to its bargaining power. So what we have obtained is a single-tier frontier model (which is yet to be augmented by the classical random error term). In retrospect, we should perhaps have anticipated that because by introducing productivity uncertainty the frontier from the part of the firm becomes equal to the expected productivity, which at the same time is the systematic component of the regression. We note that for this single-tier model, we also need to take into account the statistical dependence between the observed systematic component $\mu(\mathbf{x})$ and the single one-sided error component (which includes $\mu(\mathbf{x})$ ).

A simpler formulation could by-pass the artificial double appearance of $\mu(\mathbf{x})$ in eq. [2.28] by re-writing it as

$$
\begin{equation*}
\omega=\eta \mu(\mathbf{x})+(1-\eta) \underline{\omega} \tag{2.29}
\end{equation*}
$$

which shows that $\mu(\mathbf{x})$ is the reservation wage of the firm. The problem in implementing [2.29] is that $\eta$ is unobservable, and that realistically it differs per transaction. So [2.29] , after decomposing $\mu(\mathbf{x})$ in a regressor-format, has coefficients that are scaled by the unknown $\eta$ and are also varying. At best, it would require panel data to be estimated. And in any case the point remains that the specific attempt to develop a 2TSF bargaining framework based on Pissarides' model with uncertain productivity, is ultimately unsuccessful.

But all is not lost. In chapter 5 we formulate a proper 2TSF Nash bargaining framework, by arguing that the participants come to the negotiation table guided primarily by targets they have set, rather than by reservation thresholds.

## V. The "latent variables" argument for the 2TSF model: Economic transactions as Tug-of-War games, and beyond.

We close this chapter by a general argument in favor of the implementation of a 2TSF model.

It is certainly preferable when one can obtain the 2TSF specification starting from a structural model -it provides more depth to the whole endeavor. But the need for a model such as the 2TSF one arose from the realization (or the convincing argument) that unobserved influences exist that impact the observed outcomes. We may be able to include these influences in an underlying model, and it is then all for the better. But even if we cannot, we have the right, or rather, the obligation, by laying down economic and behavioral arguments, to include these influences in our regression specifications, if we can obtain some quantitative assessment of them in this way. Essentially, we are identifying "latent variables" -and the 2TSF model allows us to handle not one by two of them (as long as they impact the dependent variable in opposite directions). Even if we assume that these latent variables are not correlated with the regressors (and so ignoring them would not affect the coefficient estimates of the regression apart from the constant term), still, since we do believe that they affect the outcome, it is of value to find a way to assess their impact too.

The critical conceptual condition in order to apply the 2TSF model is the ability to obtain an observable systematic component which reflects the determination of the dependent variable absent the unobservable competing forces. This was the case in the Polachek \& Yoon (1987) paper, where we could have a functional specification for full-information labor supply and demand, and then two shifters which took on the roles of the opposing one-sided incomplete-information components affecting the wage in the 2TSF specification. This was the case in the '"reservation price" framework of Gaynor \& Polachek (1994), where the systematic component of the 2TSF specification is the arithmetic average of the systematic parts of the two reservation price equations. This was the case with the "hedonic analysis" framework of Kumbhakar \& Parmeter (2010), where we could obtain a functional specification for the full-information determination of the house selling price. And this will be the case in the "target-wage Nash bargaining" framework we will develop in chapter 5,
where the systematic part for the wage is the expected value of the worker output under incomplete common/symmetric information, and the opposing influences come from the systematic mark-up on it from the worker side and the systematic mark-down from it from the employer side, that in turn are due to the existence of private information on both sides.

Given this, introducing opposing forces is akin to model economic transactions as Tug-of-War games where the two teams employ strengths hidden from the observer, and the 2TSF model is employed as a remedy for this "hidden strengths/latent variables" situation. This has been essentially the approach taken by many papers we have reviewed in the previous chapter, without invoking any other structural reason for the emergence of the 2TSF specification. And we cannot resist the temptation to mention again the papers of Groot \& van den Brink (2007) and Poggi (2010) that built a 2TSF model around the same decisionmaking unit: essentially the researchers here exploited the formidable ability of a human being to play Tug-of-War with itself. --

## Chapter 3

## Independence and Exogeneity

This chapter is the first to treat the statistical and econometric tools needed to estimate 2TSF models. We start by presenting some additional results on the benchmark 2TSF Exponential distributional specification. We then move on to present an alternative 2TSF specification, the Halfnormal one, where the two one-sided components are assumed to follow Half-normal distributions. To prepare for the next 2TSF specification, in section III we extend and enhance the Corrected OLS/Method of Moments estimation strategy (COLS/MM) to accommodate specifications where the composite error density is not in closed form. Then we proceed to present the semi-Gamma specification, where one of the non-negative error components is assumed to follow a Gamma distribution while the other obeys an Exponential law. This is a specification that requires the COLS/MM estimation approach developed in the previous section. Finally, we present the Generalized Exponential 2TSF specification, where both one-sided error components are assumed to follow a specific incarnation of this distribution, motivated by a desie to have strictly positive modes in the marginal distributions. The chapter contains two empirical applications. In the first we contrast the Exponential and Half-normal 2TSF specifications on the same data set. In the second, using a different data set we implement the semi-Gamma 2TSF specification with the COLS/MM estimator.

## Introduction.

Ever since Polachek \& Yoon (1987) introduced the two-tier stochastic frontier model, there has been only one distributional specification available: the Exponential one, where the two one-sided error components are assumed to follow Exponential distributions.

In this chapter we enrich this benchmark specification with some new results, and we present three new specifications under independence and regressor exogeneity, in order to enhance the tools available to the applied researchers.

The first specification is the Half-normal one, where the two non-negative error components are assumed to follow the Half-normal distribution. We developed this specification because it is also widely used in the single-tier SF models. ${ }^{1}$ While experience from these models has shown that the choice between the Exponential and the Half-normal specification does not have much effect on the estimates of the regressor coefficients, it may

[^8]impact visibly individual measures. We will explore this in a comparative empirical application in the 2TSF context.

Another qualitative similarity between the Exponential and the Half-normal distribution is that they both have their mode at zero and a monotonically declining density. From an economic point of view, this implies an assumption that the phenomenon that is modeled takes values closer to zero with higher probability than values away from it. In the single-tier SF model one could argue that such could be the case since pressures from competition and purposeful behavior are forces that make inefficiency more likely to materialize nearer zero than away from it. ${ }^{2}$

The same can be said for the 2TSF models when the one-sided error components represent incomplete information or other phenomena whose effects economic agents are expected to try to contain and minimize. But there may be cases (and indeed, we develop such a model in a later chapter), where this assumption is rather difficult to defend, and one would expect that the mode of the distribution is away from zero.

To respond to that case also, we first present the semi-Gamma distributional specification, where one of the two one-sided error components follows a Gamma distribution, while the other follows the Exponential law. This specification leads to a nonclosed form density to be estimated, and so we present also a Corrected OLS/Method of Moments estimator to handle this case. Finally, we develop the Generalized Exponential specification, where both one-sided error components follow a distribution that has its mode away from zero.

## I. The benchmark 2TSF Exponential specification.

The 2TSF Exponential specification was first presented by Polachek and Yoon (1987)³, and was refreshed and elaborated upon by Kumbhakar and Parmeter (2009). For

[^9]completeness we present here the expressions that characterize the specification, along with some new results. The assumptions made on the 2TSF composite error term $\varepsilon=v+w-u$ are
$$
v \sim N\left(0, \sigma_{v}^{2}\right), w \sim \operatorname{Exp}\left(\sigma_{w}\right), \quad u \sim \operatorname{Exp}\left(\sigma_{u}\right),
$$
where $\sigma_{w}, \sigma_{u}$ are scale parameters of Exponential distributions, and so they equal their mean and their standard deviation at the same time. The three components are assumed to be jointly independent.

The density of the composite error term is then

$$
\begin{equation*}
f_{\varepsilon}(\varepsilon)=\frac{\exp \left\{a_{1}\right\} \Phi\left(b_{1}\right)+\exp \left\{a_{2}\right\} \Phi\left(b_{2}\right)}{\sigma_{w}+\sigma_{u}} \tag{3.1}
\end{equation*}
$$

where $\Phi(\cdot)$ is the standard normal distribution function and where

$$
\begin{equation*}
a_{1}=\frac{\sigma_{v}^{2}}{2 \sigma_{u}^{2}}+\frac{\varepsilon}{\sigma_{u}}, \quad b_{1}=-\left(\frac{\varepsilon}{\sigma_{v}}+\frac{\sigma_{v}}{\sigma_{u}}\right), \quad a_{2}=\frac{\sigma_{v}^{2}}{2 \sigma_{w}^{2}}-\frac{\varepsilon}{\sigma_{w}}, \quad b_{2}=\frac{\varepsilon}{\sigma_{v}}-\frac{\sigma_{v}}{\sigma_{w}} . \tag{3.2}
\end{equation*}
$$

The distribution function of $\varepsilon$ is

$$
\begin{equation*}
F_{\varepsilon}(\varepsilon)=\Phi\left(\frac{\varepsilon}{\sigma_{v}}\right)+\frac{\sigma_{u}}{\sigma_{w}+\sigma_{u}} \exp \left\{a_{1}\right\} \Phi\left(b_{1}\right)-\frac{\sigma_{w}}{\sigma_{w}+\sigma_{u}} \exp \left\{a_{2}\right\} \Phi\left(b_{2}\right) \tag{3.3}
\end{equation*}
$$

The distribution function can be used in order to estimate probabilities related to the composite error term, but also in a Copula model when one wants to account for possible regressor endogeneity without using instrumental variables (see chapter 4).

## I.1. Measures of one-sided error components at sample and observation level.

The estimation procedure itself will provide sample averages either directly or through the moment generating function of the Exponential distribution, depending also on the
structure of specification, namely whether the dependent variable is expressed in levels or in logarithmic form.

Jondrow, Lovell, Materov \& Schmidt (1982) have introduced a method to derive also observation-specific estimates of these effects, an approach which was extended for the 2TSF model by Kumbhakar \& Parmeter $(2009,2010)$ and where the reader can find detailed discussions on the subject. This method requires that we calculate the distribution of each one-sided error term conditional on the composite error, then calculate the relevant conditional expected values (using the residuals of the estimation), and then use these expected values as estimates of the one-sided terms for each transaction. We will call this approach the JLMS approach from here on.

The conditional distributions are

$$
\begin{align*}
& f(w \mid \varepsilon)=\frac{\lambda \exp \{-\lambda w\} \Phi\left(w / \sigma_{v}+b_{1}\right)}{\chi_{2}},  \tag{3.4}\\
& f(u \mid \varepsilon)=\frac{\lambda \exp \{-\lambda u\} \Phi\left(w / \sigma_{v}+b_{2}\right)}{\chi_{1}}, \tag{3.5}
\end{align*}
$$

where $\lambda=\sigma_{w}^{-1}+\sigma_{u}^{-1}, \quad \chi_{1}=\Phi\left(b_{2}\right)+\exp \left\{a_{1}-a_{2}\right\} \Phi\left(b_{1}\right), \quad \chi_{2}=\exp \left\{a_{2}-a_{1}\right\} \chi_{1}$.

## I.1.1. Specification in levels.

If in the regression equation the dependent variable is specified in levels, the estimated means of the two one-sided error components are provided directly from the estimation procedure.

The JLMS measures are as follows:

$$
\begin{align*}
& E\left(w_{i} \mid \varepsilon_{i}\right)=\frac{1}{\lambda}+\frac{\sigma_{v}\left[\phi\left(b_{2 i}\right)+b_{2 i} \Phi\left(b_{2 i}\right)\right]}{\chi_{1 i}},  \tag{3.6}\\
& E\left(u_{i} \mid \varepsilon_{i}\right)=\frac{1}{\lambda}+\frac{\sigma_{v}\left[\phi\left(b_{1 i}\right)+b_{1 i} \Phi\left(b_{1 i}\right)\right]}{\chi_{2 i}}, \tag{3.7}
\end{align*}
$$

where $\phi(\cdot)$ is the standard normal density.

## I.1.2. Logarithmic Specification.

If the dependent variable enters the regression in logarithms (which covers both the $\log -\log$ and the semi-log specification), we need to consider the expected values of the exponentiated variables. At the sample-level, these are calculated immediately from the moment generating function of the Exponential distribution:

$$
\begin{equation*}
E(\exp \{ \pm w\})=\frac{1}{1 \mp \sigma_{w}}, \quad E(\exp \{ \pm u\})=\frac{1}{1 \mp \sigma_{u}} \tag{3.8}
\end{equation*}
$$

For the expected value of the positive-exponent cases to exist, we require that the scale parameter is smaller than unity (which is the observed empirical regularity).

The corresponding JLMS measures are

$$
\begin{align*}
& E\left(e^{w_{i}} \mid \varepsilon_{i}\right)=\frac{\lambda}{(\lambda-1) \chi_{2 i}}\left[\Phi\left(b_{1 i}\right)+\exp \left\{\frac{1}{2}\left[\left(b_{2 i}+\sigma_{v}\right)^{2}-b_{1 i}^{2}\right]\right\} \Phi\left(b_{2 i}+\sigma_{v}\right)\right],  \tag{3.9}\\
& E\left(e^{-w_{i}} \mid \varepsilon_{i}\right)=\frac{\lambda}{(1+\lambda) \chi_{2 i}}\left[\Phi\left(b_{1 i}\right)+\exp \left\{a_{2 i}-a_{1 i}-b_{2 i} \sigma_{v}+0.5 \sigma_{v}^{2}\right\} \Phi\left(b_{2 i}-\sigma_{v}\right)\right],  \tag{3.10}\\
& E\left(e^{-u_{i}} \mid \varepsilon_{i}\right)=\frac{\lambda}{(1+\lambda) \chi_{1 i}}\left[\Phi\left(b_{2 i}\right)+\exp \left\{a_{1 i}-a_{2 i}-b_{1 i} \sigma_{v}+0.5 \sigma_{v}^{2}\right\} \Phi\left(b_{1 i}-\sigma_{v}\right)\right] . \tag{3.11}
\end{align*}
$$

Here and in subsequent models, we do not present $E\left(e^{u_{i}} \mid \varepsilon_{i}\right)$, because it does not appear in any meaningful measure of interest.

The net multiplicative effect (net mark-up) on the dependent variable is also important. Although, in the specification we examine, the one-sided error components $w, u$ are assumed independent, the correct conditional multiplicative effect is not
$E\left(e^{w_{i}} \mid \varepsilon_{i}\right) E\left(e^{-u_{i}} \mid \varepsilon_{i}\right)-1$, because $w$ and $u$ are not independent conditional on $\varepsilon$. We obtained:
$E\left(e^{w_{i}} e^{-u_{i}} \mid \varepsilon_{i}\right)=\frac{\exp \left\{\left(1+\sigma_{u}\right)\left(a_{1 i}+\frac{\sigma_{v}^{2}}{2 \sigma_{u}}\right)\right\} \Phi\left(b_{1 i}-\sigma_{v}\right)+\exp \left\{\left(1-\sigma_{w}\right)\left(a_{2 i}-\frac{\sigma_{v}^{2}}{2 \sigma_{w}}\right)\right\} \Phi\left(b_{2 i}+\sigma_{v}\right)}{\exp \left\{a_{1 i}\right\} \Phi\left(b_{1 i}\right)+\exp \left\{a_{2 i}\right\} \Phi\left(b_{2 i}\right)}$.

## I.1.3. Probabilities.

Finally we can calculate the probability that $w>u$. We have at the sample level

$$
\begin{equation*}
\operatorname{Pr}(w>u)=\frac{\sigma_{w}}{\sigma_{w}+\sigma_{u}}, \tag{3.13}
\end{equation*}
$$

and at the observation level

$$
\begin{equation*}
\operatorname{Pr}\left(w_{i}>u_{i} \mid \varepsilon_{i}\right)=\frac{\exp \left\{a_{2 i}\right\} \Phi\left(b_{2 i}\right)}{\exp \left\{a_{1 i}\right\} \Phi\left(b_{1 i}\right)+\exp \left\{a_{2 i}\right\} \Phi\left(b_{2 i}\right)} \tag{3.14}
\end{equation*}
$$

From the above expressions, our contribution is equations [3.3], [3.9], [3.12], [3.13], [3.14]. The rest can be found in Kumbhakar and Parmeter (2009). ${ }^{4}$

## I.2. Skewness and Excess Kurtosis.

Due to its two-sided character, the 2TSF Exponential error may exhibit positive or negative skewness, or even be symmetric. What is important to note is that here the skewness has necessarily the same sign as the difference of the means of the two one-sided terms (see Technical Appendix). This implies that if the residuals exhibit, say, negative skewness, $\operatorname{sign}\left\{\gamma_{1}(\varepsilon)\right\}<0$, we will necessarily obtain $\hat{E}(\varepsilon)=\hat{E}(w)-\hat{E}(u)<0$ also. This is a possibly artificial restriction on the data, because, as we will see in chapter 4, if we allow for dependence between the two one-sided error components this relation may not hold.

[^10]Namely we may have $\operatorname{sign}\left\{\gamma_{1}(\varepsilon)\right\}>0$ but $\hat{E}(\varepsilon)=\hat{E}(w)-\hat{E}(u)<0$, or any other combination. So if we ignore such dependence while it exists, we may obtain artificial estimation results that are the opposite of what exists in the data.

Regarding Excess Kurtosis, we show in the Technical Appendix that the 2TSF Exponential density exhibits positive excess kurtosis, when symmetric. This means that even if the residuals from an initial regression on a data set appear to be symmetric, still, the existence of excess kurtosis is evidence that a 2TSF model may be appropriate.

## II. The 2TSF Half-normal specification. ${ }^{5}$

We consider a single-equation cross-sectional regression model, linear in the parameters,

$$
\begin{equation*}
\mathbf{y}=X \boldsymbol{\beta}+\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}=\mathbf{v}+\mathbf{w}-\mathbf{u}, \tag{3.15}
\end{equation*}
$$

where $\mathbf{y}$ is a $n \times 1$ vector containing observations of the dependent variable, usually in levels or in logs, X is a $n \times K$ matrix of full column rank containing observations of the regressors (including a constant term), $\boldsymbol{\beta}=\left(\beta_{0}, \beta\right)^{\prime}$ and $\boldsymbol{\varepsilon}=\mathbf{v}+\mathbf{w}-\mathbf{u}$ is the $n \times 1$ vector of the composite error term. The regressors are assumed strictly exogenous, in the sense of being independent in probability from the error term. The components of the error term are assumed jointly independent for each cross-section, and identically and independently distributed over cross-sections. Their distributions are assumed to be

$$
\begin{equation*}
v \sim N\left(0, \sigma_{v}^{2}\right), w \sim H N\left(\sigma_{w}\right), \quad u \sim H N\left(\sigma_{u}\right), \tag{3.16}
\end{equation*}
$$

where $\sigma_{w}, \sigma_{u}$ are the standard deviations of the symmetric zero-mean normals of which the two Half-normals are their absolute values.

[^11]Due to the independence assumption, the mean and variance of the composite error term are immediately obtained as

$$
\begin{equation*}
E(\varepsilon)=\sqrt{\frac{2}{\pi}}\left(\sigma_{w}-\sigma_{u}\right), \quad \operatorname{Var}(\varepsilon) \equiv \sigma_{\varepsilon}^{2}=\sigma_{v}^{2}+\left(1-\frac{2}{\pi}\right)\left(\sigma_{w}^{2}+\sigma_{u}^{2}\right) . \tag{3.17}
\end{equation*}
$$

Stochastic frontier models are usually reparametrized, and in our case the most convenient reparametrization proved to be the following:

$$
\theta_{1} \equiv \frac{\sigma_{w}}{\sigma_{v}}, \theta_{2} \equiv \frac{\sigma_{u}}{\sigma_{v}}, s \equiv \sqrt{\sigma_{v}^{2}+\sigma_{w}^{2}+\sigma_{u}^{2}}=\sigma_{v} \sqrt{1+\theta_{1}^{2}+\theta_{2}^{2}},
$$

while we use for compactness the following shorthands:

$$
\omega_{1} \equiv \frac{s \sqrt{1+\theta_{2}^{2}}}{\theta_{1}}, \omega_{2} \equiv \frac{s \sqrt{1+\theta_{1}^{2}}}{\theta_{2}}, \quad \lambda_{1} \equiv \frac{\theta_{2}}{\theta_{1}} \sqrt{1+\theta_{1}^{2}+\theta_{2}^{2}}, \quad \lambda_{2} \equiv \frac{\theta_{1}}{\theta_{2}} \sqrt{1+\theta_{1}^{2}+\theta_{2}^{2}} .
$$

With this notation, the density of $\varepsilon_{i}$ is

$$
\begin{equation*}
f_{\varepsilon}\left(\varepsilon_{i}\right)=\frac{2}{s} \phi\left(\varepsilon_{i} / s\right)\left[G_{1}\left(\varepsilon_{i} ; 0, \omega_{1},-\lambda_{1}\right)-G_{2}\left(\varepsilon_{i} ; 0, \omega_{2}, \lambda_{2}\right)\right], \tag{3.18}
\end{equation*}
$$

where $G(z$; location, scale, skew $)$ is the distribution function of a univariate Skew-normal random variable. For compactness we will denote them simply $G_{1 i}$ and $G_{2 i}$,

$$
\begin{align*}
G_{1 i} & \equiv G_{1}\left(\varepsilon_{i} ; 0, \omega_{1},-\lambda_{1}\right)=\int_{-\infty}^{\varepsilon_{i}} \frac{2}{\omega_{1}} \phi\left(t / \omega_{1}\right) \Phi\left(-\lambda_{1} t / \omega_{1}\right) d t \\
G_{2 i} & \equiv G_{2}\left(\varepsilon_{i} ; 0, \omega_{2}, \lambda_{2}\right)=\int_{-\infty}^{\varepsilon_{i}} \frac{2}{\omega_{2}} \phi\left(t / \omega_{2}\right) \Phi\left(\lambda_{2} t / \omega_{2}\right) d t \tag{3.19}
\end{align*}
$$

We will denote the corresponding densities by $g_{1 i}$ and $g_{2 i}$. We also define

$$
\psi_{1 i} \equiv \frac{g_{1 i}}{G_{1 i}-G_{2 i}}, \quad \psi_{2 i} \equiv \frac{g_{2 i}}{G_{1 i}-G_{2 i}}, \quad \psi_{i}=\psi_{1 i}-\psi_{2 i}=\frac{g_{1 i}-g_{2 i}}{G_{1 i}-G_{2 i}},
$$

and $\boldsymbol{\psi}_{1}, \boldsymbol{\psi}_{2}, \boldsymbol{\psi}$ will denote the corresponding $n \times 1$ column vectors. We note that the derived density maintains the general structure of the "skew" family distributions, namely a normal density multiplied by a "skewing function", but the term $\left(G_{1 i}-G_{2 i}\right)$, apart from ranging in the $(0,1)$ interval, does not satisfy the properties of skewing functions as defined and studied in abstract by Genton \& Loperfido (2005) and Hallin \& Ley (2012). This perhaps opens the road to yet another generalization of the "skew-density" concept, but this is out of the scope of the present work.

The distribution function of the Skew-normal has two alternative representations. The first one was given in Azallini (1985), using Owen's T-function studied and tabulated in Owen (1956),

$$
\begin{equation*}
G\left(\varepsilon_{i} ; \xi, \omega, \lambda\right)=\Phi\left(\varepsilon_{i} / \omega\right)-2 T\left(\varepsilon_{i} / \omega ; \lambda\right) \tag{3.20}
\end{equation*}
$$

The second expresses it in terms of the correlated bivariate standard Normal integral $\Phi_{2}$,

$$
\begin{equation*}
G\left(\varepsilon_{i} ; \xi, \omega, \lambda\right)=2 \Phi_{2}\left(\frac{\varepsilon_{i}-\xi}{\omega}, 0 ; \rho=\frac{-\lambda}{\sqrt{1+\lambda^{2}}}\right) \tag{3.21}
\end{equation*}
$$

This second expression is more convenient for empirical implementation, since $\Phi_{2}(\cdot)$ is much more widely available in software packages as a special function compared to Owen's $T$-function. Using this we can re-write the density of the composite error as

$$
\begin{equation*}
f_{\varepsilon}\left(\varepsilon_{i}\right)=\frac{4}{s} \phi\left(\varepsilon_{i} / s\right)\left[\Phi_{2}\left(\frac{\varepsilon_{i}}{\omega_{1}}, 0 ; \rho=\frac{\lambda_{1}}{\sqrt{1+\lambda_{1}^{2}}}\right)-\Phi_{2}\left(\frac{\varepsilon_{i}}{\omega_{2}}, 0 ; \rho=\frac{-\lambda_{2}}{\sqrt{1+\lambda_{2}^{2}}}\right)\right] . \tag{3.22}
\end{equation*}
$$

But the expression involving Owen's $T$-function is useful for certain theoretical derivations, as the ones we now turn to.

## II.1. Nesting of the single-tier model, and the symmetric case.

We next show that the density for the Half-normal 2TSF model nests the single-tier specification, and therefore, it also nests both a no-inefficiency model and a deterministic frontier model. Using Owen's $T$-function we have

$$
\begin{align*}
f_{\varepsilon}\left(\varepsilon_{i}\right)=\frac{2}{s} \phi\left(\varepsilon_{i} / s\right)\left\{\left[\Phi\left(\varepsilon_{i} / \omega_{1}\right)\right.\right. & \left.+2 T\left(\varepsilon_{i} / \omega_{1} ; \lambda_{1}\right)\right]  \tag{3.23}\\
- & {\left.\left[\Phi\left(\varepsilon_{i} / \omega_{2}\right)-2 T\left(\varepsilon_{i} / \omega_{2} ; \lambda_{2}\right)\right]\right\} }
\end{align*}
$$

where we have used the property $T(a,-b)=-T(a, b)$. Now, if $\sigma_{w}^{2} \rightarrow 0$ the positive Halfnormal vanishes and we have a single-tier stochastic "production frontier" specification where the error term is $\xi=v-u$. The parameters become

$$
s \rightarrow \sqrt{\sigma_{v}^{2}+\sigma_{u}^{2}} \equiv s_{2}, \omega_{1} \rightarrow \infty, \lambda_{1} \rightarrow \infty, \omega_{2} \rightarrow s_{2} / \theta_{2}, \lambda_{2} \rightarrow 0,
$$

and the density here becomes

$$
\begin{align*}
f_{\xi}\left(\xi_{i}\right. & \left.=v_{i}-u_{i}\right)=\frac{2}{s_{2}} \phi\left(\xi_{i} / s_{2}\right)\left\{[\Phi(0)+2 T(0, \infty)]-\left[\Phi\left(\theta_{2} \xi_{i} / s_{2}\right)-2 T\left(\theta_{2} \xi_{i} / s_{2}, 0\right)\right]\right\} \\
& =\frac{2}{s_{2}} \phi\left(\xi_{i} / s_{2}\right)\left\{\left[\frac{1}{2}+2 \frac{1}{4}\right]-\left[\Phi\left(\theta_{2} \xi_{i} / s_{2}\right)-0\right]\right\} \\
& =\frac{2}{s_{2}} \phi\left(\xi_{i} / s_{2}\right) \Phi\left(-\theta_{2} \frac{\xi_{i}}{s_{2}}\right) . \tag{3.24}
\end{align*}
$$

which is a Skew-normal density representing a zero-mean Normal minus a Halfnormal.

Likewise, if $\sigma_{u}^{2} \rightarrow 0$, the negative Half normal vanishes, the error term here becomes $t=v+w$, i.e. we have a single-tier stochastic "cost frontier" specification, the parameters become

$$
s \rightarrow \sqrt{\sigma_{w}^{2}+\sigma_{v}^{2}} \equiv s_{1}, \omega_{1} \rightarrow s_{1} / \theta_{1}, \lambda_{1} \rightarrow 0, \omega_{2} \rightarrow \infty, \lambda_{2} \rightarrow \infty,
$$

and by the same steps and the properties of the $\Phi$ and $T$ functions we arrive at

$$
\begin{equation*}
f_{t}\left(t_{i}=v_{i}+w_{i}\right)=\frac{2}{s_{1}} \phi\left(t_{i} / s_{1}\right) \Phi\left(\theta_{1} \frac{t_{i}}{s_{1}}\right), \tag{3.25}
\end{equation*}
$$

which is a Skew normal density representing a normal plus a Half normal random variable. ${ }^{6}$
Finally, if $\sigma_{w}^{2}=\sigma_{u}^{2}=\sigma^{2}$ the density becomes symmetric around zero, and can be compactly written as

$$
\begin{equation*}
f_{\varepsilon}\left(\left.\varepsilon_{i}\right|_{\sigma_{w}=\sigma_{u}=\sigma}\right)=\frac{8}{s} \phi\left(\varepsilon_{i} / s\right) T\left(\varepsilon_{i} / \omega_{s} ; \lambda_{s}\right), \quad \omega_{s}=s \sqrt{1+\sigma_{v}^{2} / \sigma^{2}}, \quad \lambda_{s}=s / \sigma_{v} . \tag{3.26}
\end{equation*}
$$

## II.2. Measuring the costs of inefficiency.

To focus ideas in this section we will treat the error terms as reflecting the effects of incomplete information in a buy-sell transaction.

The estimation procedure itself will provide sample-average measures of the effects of informational inefficiencies on observed transactions. The conditional distributions needed to obtain the JLMS measures are:

For the one-sided error $w$ (representing here the incomplete information effects of the buyer side of the transaction), the density of its distribution conditional on $\varepsilon$ is

[^12]\[

$$
\begin{equation*}
f_{w \mid \varepsilon}\left(w_{i} \mid \varepsilon_{i}\right)=\left(G_{1 i}-G_{2 i}\right)^{-1} \frac{2}{\omega_{w}} \phi\left(\frac{w_{i}}{\omega_{w}}-\frac{\varepsilon_{i}}{\omega_{1}}\right) \Phi\left(\lambda_{1} \frac{\left(w_{i}-\varepsilon_{i}\right)}{\omega_{1}}\right) \tag{3.27}
\end{equation*}
$$

\]

where $\omega_{w} \equiv \frac{\sigma_{w} s_{2}}{s}$, with composite coefficients as defined earlier.
For the one-sided error term $u$ (representing the seller side of the transaction), we have

$$
\begin{equation*}
f_{u \mid \varepsilon}\left(u_{i} \mid \varepsilon_{i}\right)=\left(G_{1 i}-G_{2 i}\right)^{-1} \frac{2}{\omega_{u}} \phi\left(\frac{u_{i}}{\omega_{u}}+\frac{\varepsilon_{i}}{\omega_{2}}\right) \Phi\left(\lambda_{2} \frac{\left(u_{i}+\varepsilon_{i}\right)}{\omega_{2}}\right), \tag{3.28}
\end{equation*}
$$

where $\omega_{u} \equiv \frac{\sigma_{u} s_{1}}{s}$.
We now distinguish between the specification in levels and the log-log or semi-log ones. Moreover, we distinguish between three sets of measures: average measures, transaction-specific measures, and relative individual-specific measures.

## II.2.1. Specification in levels.

If the regression equation is specified in levels, the estimated means of the two onesided error components,

$$
\hat{\mu}_{w}=(\sqrt{2 / \pi}) \hat{\sigma}_{w},-\hat{\mu}_{u}=-(\sqrt{2 / \pi}) \hat{\sigma}_{u}
$$

provide estimates of the average upward (downward) deviations from full-information price. Percentage deviations are obtained by dividing them by the estimated average fullinformation price, which is $\bar{y}-\hat{\mu}_{w}+\hat{\mu}_{u}$, where $\bar{y}$ is the sample average of the dependent variable. Summing the two gives the net effect.

Transaction-specific measures are given by $E\left(w_{i} \mid \varepsilon_{i}\right)$ (how much the full-information price on transaction $i$ was upwardly affected by the buyer's informational inefficiency) and $-E\left(u_{i} \mid \varepsilon_{i}\right)$ (how much the full-information price was downwardly affected by the seller's informational inefficiency). The net effect on transaction $i$ is their sum. In the Half-normal specification, these conditional expected values are

$$
\begin{align*}
& E\left(w_{i} \mid \varepsilon_{i}\right)=s^{2} \psi_{1 i}+\frac{\sigma_{w}^{2}}{s^{2}}\left(\varepsilon_{i}-s^{2} \psi_{i}\right)  \tag{3.29}\\
& E\left(u_{i} \mid \varepsilon_{i}\right)=s^{2} \psi_{2 i}-\frac{\sigma_{u}^{2}}{s^{2}}\left(\varepsilon_{i}-s^{2} \psi_{i}\right) . \tag{3.30}
\end{align*}
$$

Kumbhakar \& Parmeter (2009) introduced relative individual measures of the costs of inefficiency, defined as percentage distances from each-side's optimum, and conditional also on the other side's inefficiency. This means that each is measured against a different benchmark, and so they are only conditionally comparable. (see also the authors' 2010 paper where these measures are discussed). We note that the authors present percentage efficiency measures ("how much of the distance has been covered"), while we present here their complements to unity, i.e. inefficiency measures.

Denote $\bar{p}_{b}$ the reservation price of the buyer, i.e. the maximum it is willing to pay. Denote $\underline{p}_{s}$ the minimum price the seller is willing to accept, its reservation price. Then from the buyer's point of view, the realized price can be decomposed into $p=\underline{p}_{s}+w$.

The part of the realized wage due to the buyer's incomplete information is $w$, and in percentage terms of the whole cost, $w / p$. Since $w$ is unobservable we use its conditional expected value to arrive at (since, in general notation $p=y$ )

$$
\begin{equation*}
\text { Inefficiency cost to buyer }=\frac{E\left(w_{i} \mid \varepsilon_{i}\right)}{y_{i}} \text {. } \tag{3.31}
\end{equation*}
$$

From the seller's point of view the realized price can be decomposed into $p=\bar{p}_{b}-u$. The price that the seller would have received if they had full information would be $\bar{p}_{b}=p+u$. The percentage loss relative to this best-case scenario (without the minus sign) is

$$
\begin{equation*}
\text { Inefficiency cost to seller } \frac{\bar{p}_{b}-p}{\bar{p}_{b}}=\frac{u}{p+u} \rightarrow \frac{E\left(u_{i} \mid \varepsilon_{i}\right)}{y_{i}+E\left(u_{i} \mid \varepsilon_{i}\right)} . \tag{3.32}
\end{equation*}
$$

## II.2.2. Logarithmic Specification.

If the dependent variable appears in its logarithm, then the sample-average gross mark-up (mark-down) effect of incomplete information can be estimated by $\hat{E}\left(e^{w}\right)$ and by $\hat{E}\left(e^{-u}\right)$ correspondingly, while the combined net effect is multiplicative: $\hat{E}\left(e^{w}\right) \hat{E}\left(e^{-u}\right)-1$. These unconditional expected values can be easily obtained from the moment generating function of a Half-normal random variable,

$$
\begin{equation*}
E\left(e^{t x}\right)=M G F_{H N}(t)=2 \exp \left\{\frac{1}{2} \sigma^{2} t^{2}\right\} \Phi(\sigma t) . \tag{3.33}
\end{equation*}
$$

Their theoretical values are:

$$
\begin{align*}
& E\left(e^{w}\right)=2 \exp \left\{\frac{1}{2} \sigma_{w}^{2}\right\} \Phi\left(\sigma_{w}\right)=\sqrt{\frac{2}{\pi}} \frac{\Phi\left(\sigma_{w}\right)}{\phi\left(\sigma_{w}\right)},  \tag{3.34}\\
& E\left(e^{-u}\right)=2 \exp \left\{\frac{1}{2} \sigma_{u}^{2}\right\} \Phi\left(-\sigma_{u}\right)=\sqrt{\frac{2}{\pi}} \frac{\Phi\left(-\sigma_{u}\right)}{\phi\left(\sigma_{u}\right)} . \tag{3.35}
\end{align*}
$$

Transaction-specific effects (mark-up and mark-down on full information price) are analogously given by $\hat{E}\left(e^{w_{i}} \mid \varepsilon_{i}\right)$ (buyer's side) and by $\hat{E}\left(e^{-u_{i}} \mid \varepsilon_{i}\right)$ (seller's side). The net effect is given by $\hat{E}\left(e^{w_{i}} e^{-u_{i}} \mid \varepsilon_{i}\right)-1$. As before, the variables $w_{i}, u_{i}$ are not independent when conditioned on $\varepsilon_{i}$, so setting $z_{i}=w_{i}-u_{i}$ we need to calculate $E\left(e^{w_{i}} e^{-u_{i}} \mid \varepsilon_{i}\right)-1=E\left(e^{z_{i}} \mid \varepsilon_{i}\right)-1$.

Turning to the relative individual measures of the costs of inefficiency in each transaction $i$, they can be estimated by $1-\hat{E}\left(e^{-w_{i}} \mid \varepsilon_{i}\right)$ for the buyer side, and by $1-\hat{E}\left(e^{-u_{i}} \mid \varepsilon_{i}\right)$ for the seller side.

The conditional expected values needed for the above measures are:

$$
\begin{align*}
& E\left(e^{-w} \mid \varepsilon_{i}\right)=2\left(G_{1 i}-G_{2 i}\right)^{-1} \exp \left\{\frac{1}{2} \omega_{w}^{2}-\frac{\omega_{w}}{\omega_{1}} \varepsilon_{i}\right\}\left[\Phi\left(-\frac{\left(\varepsilon_{i}+\sigma_{w}^{2}\right)}{\omega_{2}}\right)\right.  \tag{3.36}\\
&\left.-\Phi_{2}\left(-\frac{\left(\varepsilon_{i}+\sigma_{w}^{2}\right)}{\omega_{2}}, \omega_{w}-\frac{\varepsilon_{i}}{\omega_{1}} ; \rho=\frac{-\sigma_{w} \sigma_{u}}{s_{1} s_{2}}\right)\right], \\
& E\left(e^{w} \mid \varepsilon_{i}\right)=2\left(G_{1 i}-G_{2 i}\right)^{-1} \exp \left\{\frac{1}{2} \omega_{w}^{2}+\frac{\omega_{w}}{\omega_{1}} \varepsilon_{i}\right\}\left[\Phi\left(-\frac{\left(\varepsilon_{i}-\sigma_{w}^{2}\right)}{\omega_{2}}\right)\right. \\
&\left.-\Phi_{2}\left(-\frac{\left(\varepsilon_{i}-\sigma_{w}^{2}\right)}{\omega_{2}},-\left(\omega_{w}+\frac{\varepsilon_{i}}{\omega_{1}}\right) ; \rho=\frac{-\sigma_{w} \sigma_{u}}{s_{1} s_{2}}\right)\right]  \tag{3.37}\\
& E\left(e^{-u} \mid \varepsilon_{i}\right)=2\left(G_{1 i}-G_{2 i}\right)^{-1} \exp \left\{\frac{1}{2} \omega_{u}^{2}+\frac{\omega_{u}}{\omega_{2}} \varepsilon_{i}\right\}\left[\Phi\left(\frac{\varepsilon_{i}-\sigma_{u}^{2}}{\omega_{1}}\right)\right. \\
&\left.-\Phi_{2}\left(\frac{\varepsilon_{i}-\sigma_{u}^{2}}{\omega_{1}}, \frac{\varepsilon_{i}}{\omega_{2}}+\omega_{u} ; \rho=\frac{-\sigma_{w} \sigma_{u}}{s_{1} s_{2}}\right)\right]  \tag{3.38}\\
& E\left(e^{w_{i}} e^{-u_{i}} \mid \varepsilon_{i}\right)=\exp \left\{\frac{\left(\sigma_{w}^{2}+\sigma_{u}^{2}\right)}{s^{2}}\left(\frac{\sigma_{v}^{2}}{2}+\varepsilon_{i}\right)\right\} \frac{G\left(\varepsilon_{i} ;-\sigma_{v}^{2}, \omega_{1},-\lambda_{1}\right)-G\left(\varepsilon_{i} ;-\sigma_{v}^{2}, \omega_{2}, \lambda_{2}\right)}{G_{1}\left(\varepsilon_{i} ; 0, \omega_{1},-\lambda_{1}\right)-G_{2}\left(\varepsilon_{i} ; 0, \omega_{2}, \lambda_{2}\right)} \tag{3.39}
\end{align*}
$$

or, in terms of the bivariate standard normal integral,

$$
\begin{align*}
E(\exp \{w-u\} \mid \varepsilon)= & \exp \left\{\frac{\left(\sigma_{w}^{2}+\sigma_{u}^{2}\right)}{s^{2}}\left(\frac{\sigma_{v}^{2}}{2}+\varepsilon\right)\right\} \times \\
& \times \frac{\Phi_{2}\left(\frac{\varepsilon_{i}+\sigma_{v}^{2}}{\omega_{1}}, 0 ; \rho=\frac{\lambda_{1}}{\sqrt{1+\lambda_{1}^{2}}}\right)-\Phi_{2}\left(\frac{\varepsilon_{i}+\sigma_{v}^{2}}{\omega_{2}}, 0 ; \rho=\frac{-\lambda_{2}}{\sqrt{1+\lambda_{2}^{2}}}\right)}{\Phi_{2}\left(\frac{\varepsilon_{i}}{\omega_{1}}, 0 ; \rho=\frac{\lambda_{1}}{\sqrt{1+\lambda_{1}^{2}}}\right)-\Phi_{2}\left(\frac{\varepsilon_{i}}{\omega_{2}}, 0 ; \rho=\frac{-\lambda_{2}}{\sqrt{1+\lambda_{2}^{2}}}\right)} . \tag{3.40}
\end{align*}
$$

## II.3. Maximum Likelihood estimation.

The log-likelihood corresponding to model [3.15] and density [3.18] is

$$
\begin{equation*}
\tilde{L}(\boldsymbol{\varepsilon} \mid \mathbf{y}, \mathrm{X}, \mathbf{q})=n \ln \left(\frac{2}{\sqrt{2 \pi}}\right)-n \ln s-\frac{1}{2 s^{2}} \sum_{i=1}^{n}\left(y_{i}-\mathbf{x}_{i}^{\prime} \boldsymbol{\beta}\right)^{2}+\sum_{i=1}^{n} \ln \left(G_{1 i}-G_{2 i}\right), \tag{3.41}
\end{equation*}
$$

where $\mathbf{q}=\left(\boldsymbol{\beta}^{\prime}, s, \theta_{1}, \theta_{2}\right)^{\prime}$ and $\mathbf{x}_{i}^{\prime}=\left(x_{i 1}, \ldots, x_{i K}\right)$. Note that under the reparametrization, the scale terms $\theta_{1}, \theta_{2}$ are present only in the last term, while the skew parameters in $G_{1 i}$ and $G_{2 i}$ do not depend on $s$. In the Technical Appendix of the chapter we also include the derivatives of the log-likelihood.

## II.3.1. Concavity of the log-likelihood and asymptotic properties of the MLE.

The density $f_{\varepsilon}\left(\varepsilon_{i}\right)$ is the convolution of a Skew-normal and a Half-normal density. Skew-normal and Half-normal densities are each log-concave w.r.t to their variable, and the convolution operation preserves log-concavity (see for example theorem 1.3 in Brascamp \& Lieb 1975). So $f_{\varepsilon}\left(\varepsilon_{i}\right)$, viewed as a whole, is $\log$-concave w.r.t. $\varepsilon_{i}$. In the Technical Appendix we also show that $\left(G_{1 i}-G_{2 i}\right)$ separately is log-concave in $\varepsilon_{i}$. Although this is not a necessary condition for the log-concavity of $f_{\varepsilon}\left(\varepsilon_{i}\right)$ in $\varepsilon_{i}$ (it is only sufficient), it is needed in order to show concavity of the log-likelihood in the parameters, which we also do in the Technical Appendix. Given concavity, and the usual regularity conditions, the maximum likelihood estimator will be consistent and asymptotically normal (always under regressor exogeneity).

## II.3.2. OLS bias and orthogonality conditions characterizing the MLE.

In this subsection we determine the bias of the OLS estimator in the model, and we also derive some orthogonality conditions that spring out of maximum likelihood estimation related to the factor $\psi_{i}$ and its components, as defined earlier. Using the derivative of the log-likelihood, we have

$$
\begin{align*}
& \begin{aligned}
& \frac{\partial \tilde{L}}{\partial \boldsymbol{\beta}}= \frac{1}{s^{2}}\left(\mathrm{X}^{\prime} \mathbf{y}-\mathrm{X}^{\prime} \mathbf{X} \boldsymbol{\beta}\right)-\mathrm{X}^{\prime} \boldsymbol{\psi}=0 \Rightarrow \hat{\boldsymbol{\beta}}_{M L}=\left(\mathrm{X}^{\prime} \mathrm{X}\right)^{-1} \mathrm{X}^{\prime}\left(\mathbf{y}-\hat{s}_{M L}^{2} \hat{\boldsymbol{\psi}}\right) \\
&=\hat{\boldsymbol{\beta}}_{O L S}-\left(\mathrm{X}^{\prime} \mathrm{X}\right)^{-1} \mathrm{X}^{\prime}\left(\hat{s}_{M L}^{2} \hat{\boldsymbol{\Psi}}\right)=\boldsymbol{\beta}+\left(\mathrm{X}^{\prime} \mathrm{X}\right)^{-1} \mathrm{X}^{\prime}\left(\boldsymbol{\varepsilon}-\hat{s}_{M L}^{2} \hat{\boldsymbol{\psi}}\right) \\
& \Rightarrow \hat{\boldsymbol{\beta}}_{M L}=\boldsymbol{\beta}+\left(\mathrm{X}^{\prime} \mathbf{X}\right)^{-1} \mathrm{X}^{\prime} \boldsymbol{\varepsilon}-\left(\mathrm{X}^{\prime} \mathbf{X}\right)^{-1} \mathrm{X}^{\prime} \hat{s}_{M L}^{2} \hat{\boldsymbol{\psi}} .
\end{aligned}
\end{align*}
$$

Per assumptions, the regressors and the error term are mean-independent and so uncorrelated. But since $E(\boldsymbol{\varepsilon}) \neq \mathbf{0}$ (except in the special case where the two one-sided components have equal expected value), it follows that the regressors and the error term are not orthogonal. Since the MLE is consistent, it follows that

$$
\begin{equation*}
\operatorname{plim}\left(n^{-1} \mathrm{X}^{\prime} \boldsymbol{\varepsilon}\right)=\operatorname{plim}\left(n^{-1} \mathrm{X}^{\prime} \hat{s}_{M L}^{2} \hat{\boldsymbol{\psi}}\right) \Rightarrow E\left(\mathbf{x}_{i} \varepsilon_{i}\right)=E\left(\mathbf{x}_{i} s^{2} \psi_{i}\right) \tag{3.43}
\end{equation*}
$$

where $\psi_{i}$ is a function of the error term. Regressors are assumed strictly exogenous to the error term, $E(\boldsymbol{\varepsilon} \mid \mathbf{X})=E(\boldsymbol{\varepsilon})$ from which we obtain $E\left(\mathbf{x}_{i} \varepsilon_{i}\right)=E\left(\mathbf{x}_{i}\right) E\left(\varepsilon_{i}\right)$. Moreover, we show in the Technical Appendix that

$$
\begin{equation*}
E\left(s^{2} \psi_{i}\right)=E\left(\varepsilon_{i}\right)=\sqrt{(2 / \pi)}\left(\sigma_{w}-\sigma_{u}\right) \tag{3.44}
\end{equation*}
$$

Together with [3.43], these imply that the regressors are also uncorrelated with $\psi_{i}$,

$$
\begin{equation*}
E\left(\mathbf{x}_{i} s^{2} \psi_{i}\right)=E\left(\mathbf{x}_{i}\right) E\left(s^{2} \psi_{i}\right) \Rightarrow E\left(\mathbf{x}_{i} \psi_{i}\right)=E\left(\mathbf{x}_{i}\right) E\left(\psi_{i}\right) . \tag{3.45}
\end{equation*}
$$

So asymptotically, the bias of the OLS estimator due to not taking into account the nonzero mean of the composite error term is

$$
\operatorname{asymp~Bias}_{O L S}=\operatorname{plim} \hat{\boldsymbol{\beta}}_{O L S}-\boldsymbol{\beta}=\left[E\left(\mathbf{x}_{i} \mathbf{x}_{i}^{\prime}\right)\right]^{-1} E\left(\mathbf{x}_{i}\right) E\left(s^{2} \psi_{i}\right) .
$$

But we show in the Technical Appendix that $\left[E\left(\mathbf{x}_{i} \mathbf{x}_{i}^{\prime}\right)\right]^{-1} E\left(\mathbf{x}_{i}\right)=(1 \mathbf{0})^{\prime}$. In fact, this is an exact algebraic property that holds also if we use sample means, namely we have $\left(n^{-1} \mathrm{X}^{\prime} \mathrm{X}\right)^{-1}\left(n^{-1} \mathrm{X}^{\prime} \cdot \mathbf{1}\right)=(1 \mathbf{0})^{\prime}$, with $\mathbf{1}$ being a column vector of ones. These results verify that all the bias and inconsistency of the OLS estimator affects only the constant term:

$$
\operatorname{Bias}_{O L S}=E\left(\hat{\boldsymbol{\beta}}_{O L S}\right)-\boldsymbol{\beta}=\left[\begin{array}{c}
E\left(s^{2} \psi_{i}\right)  \tag{3.46}\\
\mathbf{0}
\end{array}\right]=\operatorname{plim} \hat{\boldsymbol{\beta}}_{O L S}-\boldsymbol{\beta} .
$$

Then it is immediate to obtain that the OLS estimator of the variance and of higher central moments is consistent:

$$
\begin{equation*}
\frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{i, O L S}^{r}=\frac{1}{n} \sum_{i=1}^{n}\left(\varepsilon_{i}-\mathbf{x}_{i}^{\prime}\left(\hat{\boldsymbol{\beta}}_{\text {OLS }}-\boldsymbol{\beta}\right)\right)^{r} \xrightarrow{p} E\left[\varepsilon_{i}-E\left(\varepsilon_{i}\right)\right]^{r}, r=2,3, \ldots \tag{3.47}
\end{equation*}
$$

This result will be used in the next section to formulate a Corrected OLS/Method of Moments estimator.

Turning to the MLE for the variance, we have

$$
\begin{align*}
& \frac{\partial \tilde{L}}{\partial s}=0 \Rightarrow \frac{1}{s^{2}} \sum_{i=1}^{n} \varepsilon_{i}^{2}-n-\sum_{i=1}^{n} \varepsilon_{i} \psi_{i}=0 \Rightarrow \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i}^{2}=s^{2}\left(1+\frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i} \psi_{i}\right) \\
& \Rightarrow \hat{s}_{M L}^{2}=\left(1+\frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{i} \hat{\psi}_{i}\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{i}^{2} \xrightarrow{p} \frac{E\left(\varepsilon^{2}\right)}{1+E\left(\varepsilon_{i} \psi_{i}\right)}, \tag{3.48}
\end{align*}
$$

due to the ergodic stationary sample. Moreover by the consistency of the MLE we have $\operatorname{plim} \hat{s}_{M L}^{2}=s^{2}$. Combining this with [3.48] and using the definition for the variance $s^{2} \equiv E\left(\varepsilon^{2}\right)-[E(\varepsilon)]^{2} \Rightarrow E\left(\varepsilon^{2}\right)=s^{2}+[E(\varepsilon)]^{2}$ we have

$$
s^{2}=\frac{E\left(\varepsilon^{2}\right)}{1+E\left(\varepsilon_{i} \psi_{i}\right)}=\frac{s^{2}+[E(\varepsilon)]^{2}}{1+E\left(\varepsilon_{i} \psi_{i}\right)} .
$$

From [3.44] we have $E\left(s^{2} \psi_{i}\right)=E\left(\varepsilon_{i}\right) \Rightarrow\left[E\left(\varepsilon_{i}\right)\right]^{2}=\left[E\left(s^{2} \psi_{i}\right)\right]^{2}$. Substituting,

$$
s^{2}=\frac{s^{2}+\left[s^{2} E\left(\psi_{i}\right)\right]^{2}}{1+E\left(\varepsilon_{i} \psi_{i}\right)}=s^{2} \frac{1+s^{2}\left[E\left(\psi_{i}\right)\right]^{2}}{1+E\left(\varepsilon_{i} \psi_{i}\right)},
$$

which implies

$$
1+E\left(\varepsilon_{i} \psi_{i}\right)=1+s^{2}\left[E\left(\psi_{i}\right)\right]^{2} \Rightarrow E\left(\varepsilon_{i} \psi_{i}\right)=s^{2}\left[E\left(\psi_{i}\right) E\left(\psi_{i}\right)\right]=E\left(s^{2} \psi_{i}\right) E\left(\psi_{i}\right)
$$

But we have $E\left(s^{2} \psi_{i}\right)=E\left(\varepsilon_{i}\right)$ so we arrive at

$$
\begin{equation*}
E\left(\varepsilon_{i} \psi_{i}\right)=E\left(\varepsilon_{i}\right) E\left(\psi_{i}\right) \tag{3.49}
\end{equation*}
$$

Therefore, $\operatorname{Cov}\left(\varepsilon_{i}, \psi_{i}\right)=0$, they are uncorrelated even though $\psi_{i}$ is a function of $\varepsilon_{i}$.
So $\psi_{i}$ is uncorrelated both with the error term and with the regressors (and at the same time it is orthogonal only to those variables that have a zero mean).

We also show in the Technical Appendix that, by working the MLE first-order conditions related to the theta parameters, another orthogonality condition obeyed by the estimator is

$$
\begin{equation*}
E\left[\left(\psi_{1 i}+\psi_{2 i}\right) \varepsilon_{i}\right]=0 \tag{3.50}
\end{equation*}
$$

These non-correlation / orthogonality conditions, allow us to alternatively implement the ML estimator as a GMM estimator.

Finally, the components of $\psi_{i}$ can be used to express the mean values of the one-sided components. Using [3.44] and the expressions [3.29] and [3.30] for the conditional means of $w$ and $u$ we have

$$
\begin{align*}
& E(w)=\sqrt{\frac{2}{\pi}} \sigma_{w}=E\left[E\left(w_{i} \mid \varepsilon_{i}\right)\right]=E\left(s^{2} \psi_{1 i}\right)+\frac{\sigma_{w}^{2}}{s^{2}} E\left(\varepsilon_{i}-s^{2} \psi_{i}\right)=s^{2} E\left(\psi_{1 i}\right),  \tag{3.51}\\
& E(u)=\sqrt{\frac{2}{\pi}} \sigma_{u}=E\left[E\left(u_{i} \mid \varepsilon_{i}\right)\right]=E\left(s^{2} \psi_{2 i}\right)-\frac{\sigma_{u}^{2}}{s^{2}} E\left(\varepsilon_{i}-s^{2} \psi_{i}\right)=s^{2} E\left(\psi_{2 i}\right) . \tag{3.52}
\end{align*}
$$

## II.4. Skewness and Excess Kurtosis.

The properties described earlier for the 2TSF Exponential specification hold here too: the sign of skewness of the 2TSF Half-normal density will necessarily be the same as that of $\sigma_{w}-\sigma_{u}$. Also, the symmetric distribution [3.26] exhibits positive excess kurtosis. Both results are proven in the Technical Appendix.

## II.5. Empirical Application \#1: contrasting the Exponential and Half-normal

## 2TSF specifications.

For the first empirical study of this chapter we used part of an earnings panel-data set used in Cornwell \& Rupert (1988), specifically the data related to 1982, taken from the USA Panel Study of Income Dynamics. The data concerns heads of households between the ages of 24 and 71 in 1982, which have reported a positive wage in non-farm employment.

Regarding implementation of maximum likelihood, for the Half-normal specification we used the density expressed in terms of the bivariate standard normal integral (eq. [3.22]). Also, while the reparametrization applied for this specification is convenient for theoretical results, for the application we estimated the three original parameters $\sigma_{v}, \sigma_{w}, \sigma_{u}$. This has the advantage of providing directly the standard errors on the parameters of interest.

Table 1 contains the estimation results from OLS estimation as well as from the two specifications of the 2TSF model.

Table 1: Earnings equation, estimation results.

| Dependent variable: LWAGE <br> Sample size: 595 , cross-section, year: 1982 |  | 2TSF ML Estimation |  |
| :---: | :---: | :---: | :---: |
| Variable name | OLS | Exponential <br> Specification | Half-normal <br> Specification |
| Constant | 5.5901(0.1901) | 5.5990(0.1844) | 5.6103(0.19) |
| EXP | 0.0294(0.0065) | 0.0290(0.0065) | 0.0293(0.0065) |
| EXP2 | -0.0005(0.0001) | -0.0005(0.0001) | -0.0005(0.0001) |
| WKS | 0.0034(0.0027) | 0.0038(0.0025) | $0.0036(0.0025)$ |
| OCC | -0.1615(0.0369) | -0.1649(0.037) | -0.1639(0.0373) |
| IND | $0.0847(0.0292)$ | $0.0839(0.0297)$ | $0.0833(0.0306)$ |
| SOUTH | -0.0588(0.0309) | -0.0604(0.0298) | -0.0587(0.0301) |
| SMSA | 0.1662(0.0296) | 0.1659(0.0292) | $0.1657(0.0296)$ |
| MS | 0.0952(0.0489) | 0.0943(0.0458) | 0.0939(0.0469) |
| FEM | -0.3246(0.0607) | -0.3218(0.0666) | -0.3275(0.0679) |
| UNION | 0.1063(0.0317) | 0.1093(0.033) | $0.1061(0.0336)$ |
| ED | 0.0572(0.0066) | 0.0561(0.0062) | $0.0566(0.0062)$ |
| BLK | -0.1904(0.0544) | -0.1894(0.0537) | -0.1893(0.0551) |
| $\sigma_{\varepsilon}$ | 0.3256 | 0.3219 | 0.3219 |
| $\sigma_{v}$ | - | $\begin{gathered} \hline 0.2590 \\ (0.0327) \end{gathered}$ | $\begin{gathered} 0.2198 \\ (0.0951) \end{gathered}$ |
| $\sigma_{w} / \operatorname{stdev}(w) / \mu_{w}$ | - | $\begin{aligned} & 0.1316 ~ / ~ 0.1316 ~ / ~ \\ & 0.1316(0.0357) \end{aligned}$ | $\begin{aligned} & 0.2637 / 0.1589 / \\ & 0.2104(0.1116) \end{aligned}$ |
| $\sigma_{u} / \operatorname{stdev}(u) / \mu_{u}$ | - | $\begin{aligned} & \hline 0.1388 \text { / } 0.1388 / \\ & 0.1388(0.0399) \end{aligned}$ | $\begin{array}{\|cc\|} \hline 0.2876 & / 0.1734 / \\ 0.2295 & (0.1040) \\ \hline \end{array}$ |
| $\bar{R}^{2} / \log L$ | 0.4485 / -169.98 | / -168.77 | / -169.44 |
| Residuals - skewness | -0.017 | -0.015 | -0.021 |
| Residuals - ex. kurtosis | 0.339 | 0.352 | 0.342 |
| Numbers are rounded to 4-decimal digit. Asymptotic standard errors in parentheses. Standard deviations and means of the one-sided error terms for the Half-normal case are calculated based on the estimates of the scale parameters. <br> Description of variables: $($ LWAGE $)=$ Logarithm of wage, $(E X P)=$ Years of full-time work experience, $($ EXP2 $)=$ Square of EXP, $($ WKS $)=$ Weeks worked, $(O C C)=1$ if the individual is in a blue-collar occupation, (IND) $=1$ if the individual works in a manufacturing industry, $(S O U T H)=1$ if the individual resides in the South, $(S M S A)=1$ if the individual resides in a standard metropolitan statistical area, $(\mathrm{MS})=1$ if the individual is married, (FEM) $=1$ if the individual is female, $(\mathrm{UNION})=1$ if the individual's wage is set by a union contract, $(\mathrm{ED})=$ Years of education, (BLK) $=1$ if the individual is black. |  |  |  |

We see that the three specifications produce virtually identical results related to the regressor coefficients estimates. Also, the standard deviation of the estimated composite
error term is identical in the 2TSF specifications, and pretty close to the corresponding OLS estimate. In terms of the maximized value of the log-likelihood (assumed normal for the OLS case), the three models are indistinguishable, and one does not really need to run formal tests to verify that they are equivalent from that perspective.

Some results from statistical tests on the OLS residuals are as follows: We applied the Godfrey-Orme (1991) test for zero skewness in the OLS residuals, which is robust against non-normality. It gave a test-statistic value of 0.01687 . This statistic is distributed asymptotically as a chi-square, and the $p$-value is 0.9 . So there is no evidence of skewness in the error term (as it should be expected from the estimated value of the skewness coefficient).

We performed next a zero excess-kurtosis test. Under the null hypothesis of normality, the sampling distribution of the excess kurtosis coefficient is $N(0,24 / n), n$ being the size of the sample. We obtained a (chi-square) test-statistic value of 2.847 with $p$-value of 0.092 . Zero excess-kurtosis is not rejected at the $1 \%$ and $5 \%$ significance levels, but it is rejected at the $10 \%$ level: there are indications of positive excess kurtosis in the error term, which is the distinguishing feature of a 2TSF in the presence of symmetry. In all, we have the case of an essentially symmetric error term together with evidence that the effects investigated by the 2TSF specifications are nevertheless there. The near-symmetry of the error term distribution carries over in the 2TSF models, as evidenced by the close estimated values of the two scale parameters of the one-sided error components under the two specifications. Moving to the measures of the effects of inefficiency, Table 2 presents the relevant results for the sample average level.

Table 2: Sample-average effects of informational inefficiency.

| Quantity | Interpretation | Exponential <br> Specification | Half-normal <br> Specification |
| :---: | :---: | :---: | :---: |
| $\hat{E}\left(e^{w}\right)$ | Gross mark-up on perfect-information <br> price due to informational deficiencies <br> from the side of the firm (buyer) | 1.151 | 1.251 |
| $\hat{E}\left(e^{-u}\right)$ | Gross mark-down on perfect-information <br> price due to informational deficiencies <br> from the side of the employee (seller) | 0.878 | 0.806 |
| $\hat{E}\left(e^{w}\right) \hat{E}\left(e^{-u}\right)-1$ | Net effect | 0.011 | 0.008 |

We see that under the Half-normal specification, these effects are estimated as being much stronger compared to their estimates under the Exponential specification (being further away from unity), but that the net average effect on the transactions ends up virtually identical and near zero, something to be expected since the error density appears to be symmetric around zero.

In figures $1 \& 2$ we present the kernel density estimates of the individual measures of informational inefficiency. Figure 1 contains the densities for the inefficiency measure for the employers' side (buyers), $1-\hat{E}\left(e^{-w_{i}} \mid \varepsilon_{i}\right)$, while Figure 2 displays the corresponding measures for the employee's (sellers) side, $1-\hat{E}\left(e^{-u_{i}} \mid \varepsilon_{i}\right)$. In each graph we show the estimated kernel densities under the Exponential and the Half-normal specification.

Figure 1: Individual Inefficiency - Employers (Buyers).


Figure 2: Individual Inefficiency - Employees (Sellers).


We see that the estimated densities are quite different, with the Exponential specification producing a much more peaked and concentrated density, although it shares some qualitative characteristics with the Half-normal case, like the existence of a shoulder to the right of the mode.

What is interesting here to note (and this is one of the main reasons why we chose a sample with a near symmetric composite error) is that in the space of marginal distributions, the Half-normal distribution possesses less variability than the Exponential distribution, in a mean-variance comparison. Specifically, let $X$ be a Half-normal, and let $Y$ be an Exponential, having equal means:

$$
E(X)=\sqrt{\frac{2}{\pi}} \sigma_{x}=\sigma_{y}=E(Y) \Rightarrow \sigma_{x}^{2}=\frac{\pi}{2} \sigma_{y}^{2}=\frac{\pi}{2} \operatorname{Var}(Y) .
$$

Then

$$
\operatorname{Var}(X)=\left(1-\frac{2}{\pi}\right) \sigma_{x}^{2}=\left(1-\frac{2}{\pi}\right) \frac{\pi}{2} \operatorname{Var}(Y)=\left(\frac{\pi}{2}-1\right) \operatorname{Var}(Y) \approx 0.57 \operatorname{Var}(Y)
$$

This is essentially due to the fact that the Half-normal distribution "tails off" like a Normal, faster than an Exponential. From another angle, the coefficient of variation (the standard-deviation-over-mean ratio) of an Exponential random variable is equal to 1, while the coefficient of variation of a Half-normal is equal to $\sqrt{(\pi-2) / 2} \approx 0.755 \approx \sqrt{0.57}$. And yet, we see that in the 2TSF model, the Half normal specification produces higher variability as regards the distribution of individual observations. These distributions are conditional ones, and this tells us that it is not advisable to anticipate results based on the marginals, when we work with the conditional distributions.

Turning to the net effect on each transaction, Table 3 presents the statistics for the series $\hat{E}\left(e^{w_{i}-u_{i}} \mid \varepsilon_{i}\right)-1$.

Table 3: Statistics of Net effect on Transaction series.

| $\hat{E}\left(e^{w_{i}-u_{i}} \mid \varepsilon_{i}\right)-1$ | 2TSF Exp | 2TSF HN |
| :---: | ---: | ---: |
| Mean | 0.0112 | 0.0084 |
| Median | 0.0074 | 0.0029 |
| St.Dev | 0.1224 | 0.1748 |
| Skewness | 1.6229 | 0.7151 |
| Excess Kurtosis | 10.881 | 2.2464 |
| IQ Range | 0.1273 | 0.2270 |

Although the two densities have very similar location parameters like the mean and the median, they differ in their scale and shape properties: under the Exponential specification the standard deviation is 0.122 , while under the Half-normal is 0.174 , i.e. $43 \%$ higher. Also, skewness and excess kurtosis differ markedly.

But the above differences in the estimated distributions do not essentially affect the ranking of the individuals involved in these transactions (or the ranking of each transaction). Two thirds of the observations change rank, but most move up or down just one position, and only one observation changed eight places from one specification to the other, which still
is negligible given the size of the sample: to wit, Spearman's rank correlation coefficient between the rankings from the two specifications was calculated as $99.99 \%$.

How strongly do these conditional estimates of the one-sided error terms correlate with the latter? Table 4 contains the relevant estimated correlation coefficients, since in general corr $(Y, E(Y \mid W))=S D[E(Y \mid W)] / S D(Y)$ (see Technical Appendix).

Table 4: Estimated correlation between $e^{ \pm x}$ and $E\left(e^{ \pm x} \mid \varepsilon\right)$.

| Quantity | Exponential <br> Specification | Half-normal <br> Specification |
| :--- | :---: | :---: |
| $\operatorname{côrr}\left(e^{w}, E\left(e^{w} \mid \varepsilon\right)\right)=\hat{S} D\left[E\left(e^{w} \mid \varepsilon\right)\right] / \hat{S} D\left[e^{w}\right]$ | 0.511 | 0.542 |
| $\operatorname{côrr}\left(e^{-w}, E\left(e^{-w} \mid \varepsilon\right)\right)=\hat{S} D\left[E\left(e^{-w} \mid \varepsilon\right)\right] / \hat{S} D\left[e^{-w}\right]$ | 0.460 | 0.515 |
| $\operatorname{côrr}\left(e^{-u}, E\left(e^{-u} \mid \varepsilon\right)\right)=\hat{S} D\left[E\left(e^{-u} \mid \varepsilon\right)\right] / \hat{S} D\left[e^{-u}\right]$ | 0.478 | 0.557 |
| The numerator is obtained as the sample standard deviation of each series. The denominator is <br> calculated from the moment generating functions of $w$ and $u$, using the estimated scale <br> parameters. |  |  |

The strength of correlation is average in both cases, and somewhat stronger under the Half-normal specification, an opposite result to that obtained in Waldman (1984) (which was related to the single-tier model and to population values).

In all, in this case of an essentially symmetric composite error term with mild excess kurtosis, we see that the two 2TSF specifications produce very similar results, regarding regression estimates and the ranking of observations. On the other hand they visibly differ as regards the distributional properties of the estimated individual JLMS measures.

## III. The Corrected OLS/Method of Moments estimator for 2TSF models without closed-form densities.

In this section we enhance and extend the Corrected OLS/Method of Moments estimation approach (COLS/MM), in order to be able to estimate 2TSF specifications that do not posses densities in closed form. ${ }^{7}$ The general principle of "correcting/modifying" the OLS residuals was proposed early on for the estimation of deterministic frontier models (see for example Richmond 1974). Method-of-Moments estimators have been examined in a number of cases in the literature: for example, Greene (1990) applied it in the context of the NormalGamma single-tier frontier. Kopp \& Mullahy (1990) relaxed the distributional assumptions, making only an assumption on the distribution of the one-sided error term, and developed an overidentified GMM estimator. Coelli (1995) considered using the third sample central moment of the residuals in estimating the parameters of the error distribution together with OLS estimation. Chen \& Wang (2004) presented a MM estimator for single-tier SF models with measurement error. See also Greene (2008) pp. 131-132.

In a Monte Carlo study performed by Olson, Schmidt \& Waldman (1980), the COLS/MM estimator performed well compared to, and even outperformed, the maximum likelihood estimator as regards efficiency (in terms of mean-squared error), especially in smaller samples of no more than a few hundred observations. It is therefore a secure road to take, especially when the densities involved are not closed-form.

Our contribution to the approach lies in the following: first, we derive unbiased estimators for the central moments and cumulants of the error term in a regression, which we baptized the "kapa-statistics" (to juxtapose with Fisher's unbiased " $k$-statistics" for central moments from a random sample). This is important for finite-sample performance, because higher sample moments and cumulants are known to grossly underestimate the true value in absolute terms, and the higher the order of the moment, the worse the bias. In a regression

[^13]setting, the bias also increases with the number of regressors. Indicatively, from Monte Carlo simulations that we have run, for a sample of size $n=50$ the bias of the 5th sample cumulant was $-65 \%$ with 3 regressors, and $-90 \%$ with 15 regressors. For a sample of size $n=200$ the corresponding percentage biases where $-8 \%$ and $-35 \%$ respectively.

Second, by using the kapa-statistics we re-instate the Analogy Principle at the finitesample level which increases the robustness of the method. Third, we provide kapa-statistics for up to the 5th central moment and cumulant, essentially allowing for four unknown parameters of the composite error term to be estimated in this way. Fourth, we formulate the estimator as an "estimating equations" one which allows us to implement it compactly as an exactly identified GMM estimator. Fifth, we work out explicitly the asymptotics of the estimator and provide its limiting distribution. The results can also be used in single-tier SF models.

Our motivation was the use the Gamma distribution in the 2TSF specification, so as to allow to the one-sided components to have their mode away from zero. And indeed in the next section we will present the semi-Gamma 2TSF specification that leads to a density that does not have a closed form. While for such cases more sophisticated estimation methods have been proposed, like for example the Fast-Fourier Transform (see Tsionas 2012) or simulated maximum likelihood, we decided to fully develop more simple tools for wider accessibility. ${ }^{8}$

## III.1. Corrected OLS/Method of Moments estimation - general description.

We consider the same single-equation cross-sectional linear regression model of [3.15]

$$
\mathrm{y}=\mathrm{X} \boldsymbol{\beta}+\varepsilon, \quad \varepsilon=v+w-u .
$$

In section II of this chapter, we have shown that under the assumption of regressor strict exogeneity, the OLS estimator for the beta coefficients is consistent, except for the constant term that has a finite-sample and asymptotic bias equal to the expected value of the

[^14]error term. We also obtained that the OLS estimator for the 2nd and higher central moments and cumulant of the error term is also consistent, using the inconsistent series of OLS residuals (i.e the one obtained with the inconsistent estimated value for the constant term).

The above allow us to formulate the following consistent estimation strategy combining OLS and Method of Moments (MM) estimation, the COLS/MM method:

1. Estimate the model by OLS. This will provide consistent estimates for the slope coefficients, as well as consistent standard errors for them (perhaps heteroskedasticityrobust).
2. Obtain the OLS residuals and formulate an MM estimator for the composite error parameters, by equating in the sample as many moment/cumulant equations of the error term as necessary, starting from the 2nd moment, using the uncorrected OLS residuals.
3. Provide starting values to the MM estimator by solving the system of cumulant equations numerically.
4. Run the MM estimator to obtain final MM estimates of the error parameters and their standard errors. From these, calculate the mean value of the error term $\hat{E}(\varepsilon)$.
5. Correct the OLS estimate of the constant term by subtracting from it the estimated mean of the error term (if the mean is estimated as negative, this amounts to adding it): $\hat{\beta}_{0, \text { CoLS }}=\hat{\beta}_{0, \text { oLS }}-\hat{E}(\varepsilon) .{ }^{9}$ Consequently, calculate the corrected and now consistent series of the OLS residuals, $\hat{\varepsilon}_{i, \text { COLS }}=\hat{\varepsilon}_{i, O L S}+\hat{E}(\varepsilon)$ (if the mean is estimated as negative, this amounts to subtracting it).
6. Obtain average (sample-level) measures of interest using the MM estimates and the moment generating functions of the assumed distributions.
7. Obtain observation-specific measures of interest, by plugging the corrected OLS residuals and the MM estimates of the error parameters in the various expressions. If these do not have a closed form, compute them by quadrature.
[^15]
## III.2. General Expressions for the moment equations.

We provide the theoretical expressions of the moment equations by relaxing the distributional assumptions as much as possible. This has the benefit that a researcher can use the following expressions for COLS/MM estimation by assuming different distributions, as long as the resulting composite error term contains no more than four unknown parameters. Specifically, we initially make the following assumptions only:

1. We examine the random variable $\varepsilon=v+w-u$.
2. The random variable $v$ has zero mean, is symmetric around zero and so it has all odd moments equal to zero.
3. The random variables $w$ and $u$ have non-zero mean and non-zero 3 rd, 4 th and 5 th central moments.
4. The three random variables are jointly independent.

Under these assumptions, we have the following general expressions, where the tilde indicates a variable centered on its mean, $\tilde{x}=x-E(x)$, and we use also the relations between central moments and cumulants, ${ }^{10}$

$$
\begin{equation*}
E\left(\tilde{x}^{2}\right)=\kappa_{2}=\operatorname{Var}(x), \quad E\left(\tilde{x}^{3}\right)=\kappa_{3}, \quad E\left(\tilde{x}^{4}\right)=\kappa_{4}+3 \kappa_{2}^{2}, \quad E\left(\tilde{x}^{5}\right)=\kappa_{5}+10 \kappa_{3} \kappa_{2} \tag{3.53}
\end{equation*}
$$

Then:

$$
\left\{\begin{array}{l}
E\left(\tilde{\varepsilon}^{2}\right)-\left[\kappa_{2}(v)+\kappa_{2}(w)+\kappa_{2}(u)\right]=0  \tag{3.54}\\
E\left(\tilde{\varepsilon}^{3}\right)-\left[\kappa_{3}(w)-\kappa_{3}(u)\right]=0 \\
\left(E\left(\tilde{\varepsilon}^{4}\right)-3\left[E\left(\tilde{\varepsilon}^{2}\right)\right]^{2}\right)-\left[\kappa_{4}(v)+\kappa_{4}(w)+\kappa_{4}(u)\right]=0 \\
\left(E\left(\tilde{\varepsilon}^{5}\right)-10 E\left(\tilde{\varepsilon}^{3}\right) E\left(\tilde{\varepsilon}^{2}\right)\right)-\left[\kappa_{5}(w)-\kappa_{5}(u)\right]=0
\end{array}\right.
$$

or

[^16]\[

\left\{$$
\begin{array}{l}
\kappa_{2}(\varepsilon)-\left[\kappa_{2}(v)+\kappa_{2}(w)+\kappa_{2}(u)\right]=0  \tag{3.55}\\
\kappa_{3}(\varepsilon)-\left[\kappa_{3}(w)-\kappa_{3}(u)\right]=0 \\
\kappa_{4}(\varepsilon)-\left[\kappa_{4}(v)+\kappa_{4}(w)+\kappa_{4}(u)\right]=0 \\
\kappa_{5}(\varepsilon)-\left[\kappa_{5}(w)-\kappa_{5}(u)\right]=0
\end{array}
$$\right.
\]

The above can be used to implement this estimation strategy with any triplet of distributions that respect the assumptions stated in the beginning. ${ }^{11}$

If one adds the assumption that $v$ follows the Normal distribution, then we also have $\kappa_{4}(v)=0$, and we obtain a subsystem of three equations where only the parameters of the distributions of $w$ and $u$ appear, which simplifies the initial step of providing starting values for the MM estimator,

We note that, in another context, Erickson, Jiang \& Whited (2014) provide Monte Carlo evidence that cumulant estimators perform better than moment estimators when used in a regression estimation.

## III.3. Unbiased estimation of higher moments and cumulants of the error term.

In [3.55] or [3.56], the cumulants of $\tilde{\varepsilon}$ are consistently estimated by their sample analogues using the uncorrected, inconsistent OLS residuals. But as already mentioned estimation of higher moments is known to have large downward biases. The " $k$-statistics" are unbiased estimators of cumulants, for which Fisher (1930) provides the relevant expressions, but these assume that estimation is based on draws from the true distribution.

[^17]In our case, we only have estimated values, the residuals. This requires different bias correction terms.

Moreover, unbiased estimators are needed in order to be able to properly invoke the Analogy Principle in a finite sample: The system of equations in [3.55]/[3.56] will be formulated as a set of orthogonality conditions, i.e. as expected values that equal zero, and to have that in finite samples we need unbiased estimators of the cumulants. Finally, by using unbiased estimators we clear a priori one of the hurdles to obtain limiting distributional results.

We derived unbiased estimators for the central moments and the cumulants of the error term, the "kapa-statistics", and we present them below. The various bias-correction terms are expressed in terms of the "residual maker/annihilator" matrix $\mathrm{M}=I-\mathrm{P}, \mathrm{P}=\mathrm{X}\left(\mathrm{X}^{\prime} \mathrm{X}\right) \mathrm{X}^{\prime}$. We also use the Hadamard matrix product (element-by-element multiplication), $\mathbf{M}^{(2)} \equiv \mathrm{M} \circ \mathrm{M}$. In the Technical Appendix we lay out the various properties of these matrices and of the Hadamard product that we use. All kapa-statistics remain consistent estimators.

## 2nd central moment.

For the second central moment/cumulant, i.e. the variance of the error term, the unbiased estimator is very well known,

$$
\begin{equation*}
\hat{\mu}_{2}(\varepsilon)=\hat{\kappa}_{2}(\varepsilon)=\left(c_{2}\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{i, O L S}^{2}, \quad c_{2}=\frac{1}{n} \operatorname{tr}(\mathbf{M})=\frac{n-K}{n}, \tag{3.57}
\end{equation*}
$$

where $K$ is the number of regressors (including a constant term).

## 3d central moment/cumulant.

The kapa-statistic here is

$$
\begin{equation*}
\hat{\mu}_{3}(\varepsilon)=\hat{\kappa}_{3}(\varepsilon)=\left(c_{3}\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{i, O L S}^{3}, \quad c_{3}=\frac{1}{n} \operatorname{tr}\left\{\mathbf{M}^{(2)} \mathbf{M}\right\} . \tag{3.58}
\end{equation*}
$$

4th central moment and 4th cumulant.

$$
\begin{align*}
& \hat{\mu}_{4}(\varepsilon)=\frac{c_{44}-3 c_{43}}{c_{41} c_{44}-3 c_{42} c_{43}}\left[\frac{1}{n} \sum_{j=1}^{n} \hat{\varepsilon}_{j, O L S}^{4}-\frac{\left(3 c_{42}-3 c_{41}\right)}{\left(c_{44}-3 c_{43}\right)} \frac{1}{n(n-1)} \sum_{\xi \neq \nu} \hat{\varepsilon}_{\xi, O L S}^{2} \hat{\varepsilon}_{v, O L S}^{2}\right],  \tag{3.59}\\
& \hat{\kappa}_{4}(\varepsilon)=\frac{c_{44}}{c_{41} c_{44}-3 c_{42} c_{43}}\left[\frac{1}{n} \sum_{j=1}^{n} \hat{\varepsilon}_{j, O L S}^{4}-\frac{3 c_{42}}{c_{44}} \frac{1}{n(n-1)} \sum_{\xi \neq v} \hat{\varepsilon}_{\xi, O L S}^{2} \hat{\varepsilon}_{v, O L S}^{2}\right], \tag{3.60}
\end{align*}
$$

where $\sum_{\xi \neq v}$ denotes a double sum, and

$$
\begin{gather*}
c_{41}=\frac{1}{n} \operatorname{tr}\left\{\mathbf{M}^{(2)} \mathbf{M}^{(2)}\right\}, \quad c_{42}=\frac{1}{n} \operatorname{tr}\left\{\mathbf{M}^{(2)}\right\}, \quad c_{43}=\frac{\mathbf{1}^{\prime} \mathbf{M}^{(2)} \mathbf{M}^{(2)} \mathbf{1}-n c_{41}}{n(n-1)}  \tag{3.61}\\
c_{44}=\frac{(n-K)(n-K+2)-3 n c_{42}}{n(n-1)} .
\end{gather*}
$$

## 5th central moment and 5th cumulant.

$$
\begin{align*}
& \hat{\mu}_{5}(\varepsilon)=\frac{c_{54}-10 c_{53}}{c_{51} c_{54}-10 c_{53} c_{52}}\left(\frac{1}{n} \sum_{j=1}^{n} \hat{\varepsilon}_{j, O L S}^{5}-\frac{10 c_{52}-10 c_{51}}{c_{54}-10 c_{53}} \frac{1}{n(n-1)} \sum_{\xi \neq v}^{n} \hat{\varepsilon}_{\xi, O L S}^{2} \hat{\varepsilon}_{v, O L S}^{3}\right),  \tag{3.62}\\
& \hat{\kappa}_{5}(\varepsilon)=\frac{c_{54}}{c_{51} c_{54}-10 c_{53} c_{52}}\left(\frac{1}{n} \sum_{j=1}^{n} \hat{\varepsilon}_{j, O L S}^{5}-\frac{10 c_{52}}{c_{54}} \frac{1}{n(n-1)} \sum_{\xi \neq v}^{n} \hat{\varepsilon}_{\xi, O L S}^{2} \hat{\varepsilon}_{v, O L S}^{3}\right), \tag{3.63}
\end{align*}
$$

where again $\sum_{\xi \neq v}$ denotes a double sum, and

$$
\begin{aligned}
& c_{51}=\frac{\operatorname{tr}\left\{\mathbf{M}^{(3)} \mathbf{M}^{(2)}\right\}}{n}, \quad c_{52}=\frac{\mathbf{1}^{\prime} \mathbf{M}^{(3)} \mathbf{m}_{d}}{n}, \quad c_{53}=\frac{\mathbf{1}^{\prime} \mathbf{M}^{(3)} \mathbf{M}^{(2)} \mathbf{1}-n c_{51}}{n(n-1)}, \\
& c_{54}=\frac{1}{n(n-1)}\left[(n-K+6) \operatorname{tr}\left\{\mathbf{M}^{(2)} \mathbf{M}\right\}+3\left(\mathbf{1}^{\prime} \mathbf{M}^{(2)} \mathbf{M} \mathbf{m}_{d}\right)-10 n c_{52}\right] .
\end{aligned}
$$

Here $\mathbf{m}_{d}$ is a column vector holding the main diagonal terms of M .

## III.4. The Method of Moments estimator. ${ }^{12}$

The MM estimator is a case of an "estimating equations" estimator. Using the kapastatistics and the OLS residuals we bring the system of moment equations in [3.55] into a suitable vector form with the following elements:

$$
\left\{\begin{array}{l}
\hat{h}_{n, 2} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(c_{2}\right)^{-1} \hat{\varepsilon}_{i, O L S}^{2}-\left(\kappa_{2}(v)+\kappa_{2}(w)+\kappa_{2}(u)\right)\right]  \tag{3.64}\\
\hat{h}_{n, 3} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(c_{3}\right)^{-1} \hat{\varepsilon}_{i, O L S}^{3}-\left(\kappa_{3}(w)-\kappa_{3}(u)\right)\right] \\
\hat{h}_{n, 4} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(C_{4 a} \hat{\varepsilon}_{i, O L S}^{4}-\frac{C_{4 b}}{(n-1)} \hat{\varepsilon}_{i, O L S}^{2} \sum_{\xi \neq i} \hat{\varepsilon}_{\xi, O L S}^{2}\right)-\left(\kappa_{4}(v)+\kappa_{4}(w)+\kappa_{4}(u)\right)\right] \\
\hat{h}_{n, 5} \equiv \frac{1}{n} \sum_{i=1}^{n}\left[\left(C_{5 a} \hat{\varepsilon}_{i, O L S}^{5}-\frac{C_{5 b}}{(n-1)} \hat{\varepsilon}_{i, O L S}^{3} \sum_{\xi \neq i} \hat{\varepsilon}_{\xi, O L S}^{2}\right)-\left(\kappa_{5}(w)-\kappa_{5}(u)\right)\right]
\end{array}\right.
$$

Enumeration of the $h$-functions follows the order of the moment involved (this is why the subscript " 1 " is missing). Also, here $\sum_{\xi \neq i}$ is a single-sum of $n-1$ terms, and

$$
\begin{aligned}
& C_{4 a}=\frac{c_{44}}{c_{41} c_{44}-3 c_{42} c_{43}} \xrightarrow{p} 1, \quad C_{4 b}=C_{4 a} \frac{3 c_{42}}{c_{44}} \xrightarrow{p} 3, \\
& C_{5 a}=\frac{c_{54}}{c_{51} c_{54}-10 c_{53} c_{52}} \xrightarrow{p} 1, C_{5 b}=C_{5 a} \frac{10 c_{52}}{c_{54}} \xrightarrow{p} 10 .
\end{aligned}
$$

[^18]More compactly

$$
\begin{cases}\hat{h}_{2, i} \equiv f_{2 i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)-\kappa_{2}(\varepsilon ; \mathbf{q}), & \hat{h}_{n, 2}=\frac{1}{n} \sum_{i=1}^{n} \hat{h}_{2, i}  \tag{3.65}\\ \hat{h}_{3, i} \equiv f_{3 i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)-\kappa_{3}(\varepsilon ; \mathbf{q}), & \hat{h}_{n, 3}=\frac{1}{n} \sum_{i=1}^{n} \hat{h}_{3, i} \\ \hat{h}_{4, i} \equiv f_{4 i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)-\kappa_{4}(\varepsilon ; \mathbf{q}), & \hat{h}_{n, 4}=\frac{1}{n} \sum_{i=1}^{n} \hat{h}_{4, i} \\ \hat{h}_{5, i} \equiv f_{5 i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)-\kappa_{5}(\varepsilon ; \mathbf{q}), & \hat{h}_{n, 5}=\frac{1}{n} \sum_{i=1}^{n} \hat{h}_{5, i}\end{cases}
$$

The correspondence to [3.64] is obvious. Here $\mathbf{q}$ is the vector of parameters determining the cumulants of $v, w, u$, with $\mathbf{q}_{0}$ denoting the vector with the true values. Note that the unknown parameters are additively separable from the data. The $f$-functions represent here the functions of the residuals we have used (the unbiased estimators for the cumulants), and this is why they depend on the whole sample.

We define the vector $\hat{\mathbf{h}}_{N}(\mathbf{q})=\left(\hat{h}_{n, 2}, \hat{h}_{n, 3}, \hat{h}_{n, 4}, \hat{h}_{n, 5}\right)^{\prime}$, and for later use $\hat{\mathbf{h}}_{i}(\mathbf{q})=\left(\hat{h}_{2, i}, \hat{h}_{3, i}, \hat{h}_{4, i}, \hat{h}_{5, i}\right)^{\prime}$, so $\hat{\mathbf{h}}_{N}(\mathbf{q})=n^{-1} \sum_{i=1}^{n} \hat{\mathbf{h}}_{i}(\mathbf{q})$. Then, the MM estimator solves

$$
\begin{equation*}
\hat{\mathbf{q}}: \hat{\mathbf{h}}_{N}(\hat{\mathbf{q}})=0 . \tag{3.66}
\end{equation*}
$$

## III.4.1. Properties of the MM estimator.

Since we have used the unbiased kapa-statistics, we have that

$$
\begin{equation*}
E\left[\hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)\right]=\mathbf{0} \tag{3.67}
\end{equation*}
$$

i.e. not just as a limiting result, but holding also for finite samples. This makes the $\hat{h}_{n, j}$ elements of $\hat{\mathbf{h}}_{N}$ proper orthogonality conditions, which aligns with the fact that equation
[3.66] can be obtained as the first order condition for the maximization of an M-estimator. Specifically, consider the objective function of a quadratic in sample averages,

$$
\begin{equation*}
C \equiv \frac{1}{2}\left(\hat{\mathbf{h}}_{N}(\mathbf{q})\right)^{\prime}\left(\hat{\mathbf{h}}_{N}(\mathbf{q})\right)=\frac{1}{2} \sum_{j=2}^{5} \hat{h}_{n, j}^{2} . \tag{3.68}
\end{equation*}
$$

It is evident that it reaches its minimum value of zero if and only if

$$
\hat{h}_{n, j}=0 \quad j=2,3,4,5 \Leftrightarrow \hat{\mathbf{h}}_{N}(\hat{\mathbf{q}})=0 \text {, which is eq. [3.66]. }
$$

Moreover, the objective function in [3.68] reveals the MM estimator to be a Generalized Method of Moments (GMM) estimator with the identity matrix as the weights matrix. Since the system is exactly identified, the estimator is not affected by the choice of the weights matrix.

## Consistency.

Consistency of the MM estimator comes from the consistency of the $f$-functions, which has been already shown, given the ergodic stationarity of the sample, and the standard regularity conditions: due to these we have $\hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right) \xrightarrow{p} 0$ (and this is the minimum value), while $\hat{\mathbf{h}}_{N}(\hat{\mathbf{q}})=0 \quad \forall n$, which gives consistency.

## Asymptotic Normality. ${ }^{13}$

To obtain the asymptotic distribution of the MM estimator we start by performing a mean value expansion of $\hat{\mathbf{h}}_{N}(\hat{\mathbf{q}})$ around the true vector $\mathbf{q}_{0}$,

$$
\hat{\mathbf{h}}_{N}(\hat{\mathbf{q}})=\hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)+\mathrm{J}_{h}(\overline{\mathbf{q}})\left(\hat{\mathbf{q}}-\mathbf{q}_{0}\right),
$$

[^19]where $\mathbf{J}_{h}$ is the Jacobian matrix of $\hat{\mathbf{h}}_{N}(\mathbf{q})$ with respect to $\mathbf{q}$. It includes only the unknown parameters and it is here evaluated at vectors $\overline{\mathbf{q}}$ (different for each row of the matrix) inbetween $\hat{\mathbf{q}}$ and $\mathbf{q}_{0}$. Since $\hat{\mathbf{h}}_{N}(\hat{\mathbf{q}})=0$ we obtain
\[

$$
\begin{equation*}
n^{1 / 2}\left(\hat{\mathbf{q}}-\mathbf{q}_{0}\right)=-\mathbf{J}_{h}^{-1}(\overline{\mathbf{q}})\left(n^{1 / 2} \hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)\right) . \tag{3.69}
\end{equation*}
$$

\]

Certain features of [3.69] deserve comment.
First, the inverted matrix is a Jacobian, not a Hessian, which means that it won't be symmetric (it contains the derivatives of the various cumulants with respect to the unknown parameters). Its invertibility, which is a necessary condition for identification, is not an almost-universal property as it is with Hessian matrices.

Second, this matrix does not depend on the data or on the sample size, so we do not need to assume a well defined probability limit for it, since it is a matrix of constants (although we do need to make such an assumption for its estimator). All we need is invertibility which must be checked and verified per case, once we specify the distributions involved and obtain their cumulants. Essentially what is required is to show that there does not exist a permissible combination of parameter values that will make the determinant of $\mathrm{J}_{h}(\mathbf{q})$ equal to zero (or, if such a combination exists we must be prepared to assume it away in order to proceed with the model).

Given invertibility, consistency of $\hat{\mathbf{q}}$ implies that the $\overline{\mathbf{q}}$ vectors will be squeezed to the true value $\mathbf{q}_{0}$ at the limit, so we have $\mathrm{J}_{h}^{-1}(\overline{\mathbf{q}}) \xrightarrow{p} \mathrm{~J}_{h}^{-1}\left(\mathbf{q}_{0}\right)$.

The remaining issue is the limiting joint distribution of the vector

$$
\begin{equation*}
n^{1 / 2} \hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)=\left\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left[f_{j i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)-\kappa_{j}\left(\varepsilon ; \mathbf{q}_{0}\right)\right], \quad j=2,3,4,5\right\} . \tag{3.70}
\end{equation*}
$$

Due to the use of the unbiased kapa-statistics the random variables $f_{j i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)-\kappa_{j}\left(\varepsilon ; \mathbf{q}_{0}\right)$ have zero mean $\forall j, i$. Also, for each $j$, they are identically distributed over $i$, since they are identical functions of the OLS residuals (which are
identically distributed, given the original i.i.d assumption on the sample). But they are not independent, since every residual is a function of the whole sample through the OLS estimator, and also the bias-correction factors are a function of the whole regressor matrix. In fact $f_{j i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)$ are equicorrelated (per $j$ ), but the dependence weakens to extinction as $n$ increases. So each element of the vector $n^{1 / 2} \hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)$ is the scaled sum of zero-mean, identically distributed, equicorrelated but asymptotically independent random variables.

First, we consider each such element separately. The distribution of each term in the sum changes as the sample size increases, and so the convergence of the sum must be examined in the context of triangular arrays for dependent random variables. This involves the concept of mixing (see White \& Domowitz 1984 for a focused econometric introduction and application of the concept). While mixing is almost always presented in a time-series framework, it is equally applicable to a stochastic process where no natural ordering exists (see for example how the concept is defined in Zhengyan \& Chuanrong 1996, ch. 1). When the sample is cross-sectional, then mixing and the mixing coefficients should be examined for all possible permutations of the variables. In our case, since the variables in the sums we examine are identically distributed and equicorrelated, they are exchangeable and so the order that they appear in the sample does not matter. It follows that we only need to examine mixing for just one arbitrarily determined order.

Due to asymptotic joint independence, the sequences whose elements are summed satisfy any concept of mixing (like strong or uniform mixing), since as they become infinite in length, not only elements "far apart" but also elements "side-by-side" become independent, and so the mixing coefficients are in all cases zero. In other words our sequences satisfy a property much stronger than mixing. In particular then, it is an " $\ell$-mixing" sequence as the concept is defined in Withers (1981). We show in the Technical Appendix that each element of $n^{1 / 2} \hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)$ satisfies also the other sufficient conditions of Theorem 2.1/Corollary 2.1 in this paper, and so it obeys the Normal Central Limit Theorem, converging to the standard Normal distribution if scaled by its variance.

But, even though individually we have that

$$
f_{j i}\left(\mathrm{y}_{n}, \mathrm{X}_{n}\right)-\kappa_{j}\left(\varepsilon ; \mathbf{q}_{0}\right) \xrightarrow{p / d} \tilde{\varepsilon}_{i}^{j}-\mu_{j}, \quad \forall i, \quad j=2,3,4,5,
$$

it is not true that all the elements of $n^{1 / 2} \hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)$ converge in distribution to the central error sample moments centered on the true values,

$$
\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}^{2}-\mu_{2}\right), \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}^{3}-\mu_{3}\right), \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}^{4}-\mu_{4}\right), \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}^{5}-\mu_{5}\right)\right)^{\prime} .
$$

In fact only the expression for the second cumulant/variance does so,

$$
\begin{equation*}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left[\left(c_{2}\right)^{-1} \hat{\varepsilon}_{i, O L S}^{2}-\left(\kappa_{2}(v)+\kappa_{2}(w)+\kappa_{2}(u)\right)\right]-\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}^{2}-\mu_{2}\right) \longrightarrow p 0 \tag{3.71}
\end{equation*}
$$

while for example, it holds that

$$
\begin{equation*}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left[\frac{1}{c_{3}} \hat{\varepsilon}_{i, O L S}^{3}-\kappa_{3}(\varepsilon)\right]-\left\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}^{2}-\mu_{3}\right)-3\left(\frac{1}{n} \sum_{i=1}^{n} \tilde{\varepsilon}_{i}^{2} \mathbf{x}_{i}^{\prime}\right)\left[\sqrt{n}\left(\hat{\beta}_{O L S}^{*}-\beta\right)\right]\right\} \xrightarrow{p} 0 \tag{3.72}
\end{equation*}
$$

where $\hat{\beta}_{O L S}^{*}$ is the OLS estimator corrected so as to be consistent also for the constant term. So Lemma A(iv) p. 68 in Serfling (1980) that deals with the limiting distribution of the whole vector of error central moments is inapplicable here. Still, we show, using the CramérWold theorem (or "device"), that $n^{1 / 2} \hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right)$ is multivariate Normal at the limit,

$$
\begin{equation*}
n^{1 / 2} \hat{\mathbf{h}}_{N}\left(\mathbf{q}_{0}\right) \xrightarrow{d} N\left(0, \mathrm{~V}_{h}\right), \quad \mathrm{V}_{h}=\lim _{n \rightarrow \infty} E\left[\hat{\mathbf{h}}_{i}\left(\mathbf{q}_{0}\right) \hat{\mathbf{h}}_{i}\left(\mathbf{q}_{0}\right)^{\prime}\right] . \tag{3.73}
\end{equation*}
$$

It follows that the limiting distribution of $\sqrt{n}\left(\hat{\mathbf{q}}-\mathbf{q}_{0}\right)$ is (after eliminating the minus sign in [3.69] due to the zero-mean and the symmetry properties of the normal distribution)

$$
\begin{equation*}
\sqrt{n}\left(\hat{\mathbf{q}}-\mathbf{q}_{0}\right) \xrightarrow{d} N\left(0, \mathrm{~V}_{q}\right), \quad \mathrm{V}_{q}=\mathrm{J}_{h}^{-1}\left(\mathbf{q}_{0}\right) \mathrm{V}_{h}\left(\mathrm{~J}_{h}^{-1}\left(\mathbf{q}_{0}\right)\right)^{\prime}, \tag{3.74}
\end{equation*}
$$

and so the asymptotic distribution of the MM estimator is ${ }^{14}$

$$
\begin{equation*}
\hat{\mathbf{q}} \stackrel{a}{\sim} N\left(\mathbf{q}_{0}, n^{-1} \mathrm{~V}_{q}\right) . \tag{3.75}
\end{equation*}
$$

A consistent estimator of the limiting variance-covariance matrix is

$$
\begin{equation*}
\hat{\mathrm{V}}_{q}=\mathrm{J}_{h}^{-1}(\hat{\mathbf{q}})\left(\frac{1}{n} \sum_{i=1}^{n} \hat{\mathbf{h}}_{i}(\hat{\mathbf{q}}) \hat{\mathbf{h}}_{i}(\hat{\mathbf{q}})^{\prime}\right)\left(\mathrm{J}_{h}^{-1}(\hat{\mathbf{q}})\right)^{\prime} . \tag{3.76}
\end{equation*}
$$

We stress that $\mathrm{J}_{h}^{-1}(\hat{\mathbf{q}}) \neq\left(\mathrm{J}_{h}^{-1}(\hat{\mathbf{q}})\right)^{\prime}$ since the matrix is not symmetric, and that we need to divide $\hat{\mathrm{V}}_{q}$ by $n$ to obtain the standard errors of the estimates. The middle term is the outer product or BHHH estimate (after Berndt, Hall, Hall, and Hausman 1974), and is by construction heteroskedasticity-robust.

In practice, the MM estimator can be implemented in software as a GMM estimator with the Identity matrix as the weights matrix, which will automatically provide the correct variance covariance matrix $n^{-1} \hat{\mathrm{~V}}_{q}$.

We turn now to present the semi-Gamma 2TSF specification that requires the use of the COLS/MM estimator.

## IV. The semi-Gamma 2TSF specification.

In this section, we present a 2TSF distributional specification where one of the two nonnegative components is assumed to follow a Gamma distribution. The motivation for such models is the following: first, as Stevenson (1980) noted, the ubiquitous Exponential and Half-normal distributions have their mode at zero, and this imposes the assumption that the phenomenon represented by them is more likely to take values closer to zero. As we have already discussed, this is not necessarily true for all cases where a single-tier SF model may

[^20]be applied, and even more so when we have a 2TSF structure. For example, in the Nashbargaining 2TSF framework the one-sided terms include the relative bargaining power, and we don't expect its more probable values to be almost zero.

Second, as Tsionas (2012, section 7 p. 242) has found, the hypothesis of an Exponential one-sided component has been rejected in favor of a Gamma distribution (but not for the other component). So a specification with different distributions has already been supported by the data.

The assumption of a Gamma-distributed one-sided component leads to a density for the composite error term that is not in closed form, but they can be estimated by the COLS/MM estimator developed in the previous section. Observation-specific measures are also not in closed form, but having the estimated parameters and a consistent series for the residuals, we can compute them by quadrature. The interest in obtaining measures at observation-level is also the reason why we do not propose a full Gamma 2TSF model (in which case, the conditional densities would be much more complicated to calculate numerically, and we would also need the kapa-statistic for the 6th cumulant). But we provide two variants, namely one specification where the positive error component follows a Gamma distribution, and the other where the negative one does.

## IV.1. The Gamma-Exponential 2TSF specification.

## IV.1.1. The composite error density.

Here we assume that the error components obey the following laws:

$$
\begin{equation*}
v \sim N\left(0, \sigma_{v}^{2}\right), w \sim \operatorname{Gamma}(k, \theta), \quad u \sim \operatorname{Exp}\left(\sigma_{u}\right), \tag{3.77}
\end{equation*}
$$

where for the Gamma distribution we adopt the shape-scale parametrization,

$$
\begin{equation*}
f_{w}(w)=\frac{1}{\Gamma(k) \theta^{k}} w^{k-1} \exp \{-w / \theta\} . \tag{3.78}
\end{equation*}
$$

Since only one of the variables follows a two-parameter distribution, to un-clutter notation we do not index $k, \theta$ by the subscript $w$.

Due to the independence assumption, the mean and variance of the composite error term are immediately obtained as

$$
\begin{equation*}
E(\varepsilon)=k \theta-\sigma_{u}, \quad \operatorname{Var}(\varepsilon) \equiv \sigma_{\varepsilon}^{2}=\sigma_{v}^{2}+k \theta^{2}+\sigma_{u}^{2} \tag{3.79}
\end{equation*}
$$

We obtained the composite error density starting with the difference $z=w-u$ for which we have

$$
f_{Z}(z)= \begin{cases}\frac{\sigma_{u}^{k-1} \exp \left\{z / \sigma_{u}\right\}}{\left(\sigma_{u}+\theta\right)^{k}} & z \leq 0  \tag{3.80}\\ \frac{\sigma_{u}^{k-1} \exp \left\{z / \sigma_{u}\right\}}{\left(\sigma_{u}+\theta\right)^{k}}\left[1-F_{G}(z ; k, \delta)\right] & z>0\end{cases}
$$

where $F_{G}(z ; k, \delta)$ is the Gamma distribution function with shape parameter $k$ and scale parameter $\delta \equiv \sigma_{u} \theta /\left(\sigma_{u}+\theta\right)$. Consequently, the density of the composite error term is

$$
\begin{equation*}
f_{\varepsilon}(\varepsilon)=\frac{\sigma_{u}^{k-1}}{\left(\sigma_{u}+\theta\right)^{k}}\left[\exp \left\{\frac{\varepsilon}{\sigma_{u}}+\frac{\sigma_{v}^{2}}{2 \sigma_{u}^{2}}\right\}-\int_{0}^{\infty} e^{z / \sigma_{u}} \frac{1}{\sigma_{v}} \phi\left(\frac{\varepsilon-z}{\sigma_{v}}\right) F_{G}(z ; k, \delta) d z\right] . \tag{3.81}
\end{equation*}
$$

The fact that $F_{G}(z ; k, \delta)$ is bounded by unity and that $\phi\left((\varepsilon-z) / \sigma_{v}\right)$ includes $-z^{2}$ in its exponent, guarantee that the integral converges even though the power of the initial exponential term in the integrand expression is positive. Note that the variable $z$ in [3.81] is just the dummy variable of integration.

Although we will not perform maximum-likelihood estimation, this density and its numerical computation is needed for the observation-specific measures we are interested in. The model nests the 2TSF Exponential specification by setting $k=1$.

## Skewness.

The sign of skewness of this distribution depends on the parameters of the one-sided components. Specifically we have

$$
\operatorname{sign}\left\{\gamma_{1}\right\}=\operatorname{sign}\left\{\kappa_{3}\right\}=\operatorname{sign}\left\{k \theta^{3}-\sigma_{u}^{3}\right\} .
$$

This sign will not necessarily be the same as the sign of $E(w)-E(u)=k \theta-\sigma_{u}$.
If $k<1$ it may be the case that $k \theta<\sigma_{u}<k^{1 / 3} \theta$ implying that $E(w)-E(u)<0$ but $\operatorname{sign}\left\{\gamma_{1}\right\}>0$.

If $k>1$, we may have $k^{1 / 3} \theta<\sigma_{u}<k \theta$ implying that $E(w)-E(u)>0$ but $\operatorname{sign}\left\{\gamma_{1}\right\}<0$.

So by introducing the Gamma distribution in the specification, we gain additional flexibility compared to the Exponential and Half-normal 2TSF specifications.

## IV.1.2. Moment Equations for the 2TSF Gamma-Exponential model.

The expressions in the previous subsection need to be fed parameter estimates and residuals. Applying our specific distributional assumptions, the non-linear system of equations that the MM estimator must solve becomes

$$
\begin{cases}\hat{\kappa}_{2}(\varepsilon)=\sigma_{v}^{2} & +k \theta^{2}+\sigma_{u}^{2}  \tag{3.82}\\ (1 / 2) \hat{\kappa}_{3}(\varepsilon) & =k \theta^{3}-\sigma_{u}^{3} \\ (1 / 6) \hat{\kappa}_{4}(\varepsilon) & =k \theta^{4}+\sigma_{u}^{4} \\ (1 / 24) \hat{\kappa}_{5}(\varepsilon) & =k \theta^{5}-\sigma_{u}^{5}\end{cases}
$$

Setting for compactness $a_{3}=(1 / 2) \hat{\kappa}_{3}(\varepsilon), a_{4}=(1 / 6) \hat{\kappa}_{4}(\varepsilon), a_{5}=(1 / 24) \hat{\kappa}_{5}(\varepsilon)$ we focus on the $3 \times 3$ subsystem

$$
\left\{\begin{array}{l}
a_{3}-k \theta^{3}+\sigma_{u}^{3}=0  \tag{3.83}\\
a_{4}-k \theta^{4}-\sigma_{u}^{4}=0 \\
a_{5}-k \theta^{5}+\sigma_{u}^{5}=0
\end{array}\right.
$$

Manipulating,

$$
\left\{\begin{array} { l } 
{ k \theta ^ { 3 } = \sigma _ { u } ^ { 3 } + a _ { 3 } } \\
{ k \theta ^ { 3 } = \frac { a _ { 4 } - \sigma _ { u } ^ { 4 } } { \theta } } \\
{ k \theta ^ { 3 } = \frac { \sigma _ { u } ^ { 5 } + a _ { 5 } } { \theta ^ { 2 } } }
\end{array} \Rightarrow \left\{\begin{array}{l}
\theta=\frac{a_{4}-\sigma_{u}^{4}}{\sigma_{u}^{3}+a_{3}} \\
\theta=\frac{\sigma_{u}^{5}+a_{5}}{a_{4}-\sigma_{u}^{4}}
\end{array} \Rightarrow\left(a_{4}-\sigma_{u}^{4}\right)^{2}=\left(\sigma_{u}^{3}+a_{3}\right)\left(\sigma_{u}^{5}+a_{5}\right)\right.\right.
$$

which leads to an equation in only one unknown

$$
\begin{align*}
& a_{4}^{2}-2 a_{4} \sigma_{u}^{4}+\sigma_{u}^{8}=\sigma_{u}^{8}+a_{5} \sigma_{u}^{3}+a_{3} \sigma_{u}^{5}+a_{3} a_{5} \Rightarrow a_{3} \sigma_{u}^{5}+2 a_{4} \sigma_{u}^{4}+a_{5} \sigma_{u}^{3}-a_{4}^{2}+a_{3} a_{5}=0 \\
& \Rightarrow \sigma_{u}^{3}\left(a_{3} \sigma_{u}^{2}+2 a_{4} \sigma_{u}+a_{5}\right)-\left(a_{4}^{2}-a_{3} a_{5}\right)=0 \tag{3.84}
\end{align*}
$$

After running OLS estimation and obtaining the kapa-statistics, this equation can be solved by a numerical solver, or even be graphed and have its roots determined to good accuracy by a few trial and error attempts. In turn, one can check which of the admissible (strictly positive) roots of [3.84], if any, produce permissible values also for the other three parameters. These can be used as the starting values for the MM estimator, to follow the rest of the steps described in section III.1. This procedure serves also as a check on whether the specification is admissible by the data.

Finally, we show in the Technical Appendix that the determinant of the Jacobian of the full system is

$$
\begin{equation*}
|\mathrm{J}|=-2 \sigma_{v} k \theta^{6} \sigma_{u}^{2}\left(5 \sigma_{u}^{2}+8 \sigma_{u} \theta+3 \theta^{2}\right) \tag{3.85}
\end{equation*}
$$

which can be zero only if one of the parameters is zero. Now assume that in empirical estimation, any one of the parameters is estimated as very close to zero. This would not be just a rejection of the particular distributional specification, but an indication that the three component 2TSF error structure is not compatible with the data. For, if any one of the parameters are zero (even for the two-parameter Gamma distribution), it implies that the related component of the specification does not exist as a non-degenerate random variable. If the 2TSF structure is compatible with the data, then distributional misspecification alone won't result in parameter estimates being almost zero, since the estimator will try to shape the density as best it can given the specification we have imposed. Reversely, a set of estimates "reasonably" away from zero is not an indication that the particular distributional specification is correct, only that the 2TSF structure is not grossly in conflict with the data.

## IV.1.3. Individual JLMS measures.

Once we have estimated the model, we can proceed with the calculation of various measures of interest.

For the positive Gamma random variable $w$ we want to calculate

$$
E(g(w) \mid \varepsilon)=\int_{0}^{\infty} g(w) f_{w \mid \varepsilon}(w \mid \varepsilon) d w=\frac{1}{f_{\varepsilon}(\varepsilon)} \int_{0}^{\infty} g(w) f_{w \varepsilon}(w, \varepsilon) d w,
$$

where $g(w)$ is the measure of interest, $g(w)=w, \exp \{ \pm w\}$. We obtained

$$
\begin{equation*}
E(g(w) \mid \varepsilon)=\frac{\exp \left\{\frac{\varepsilon}{\sigma_{u}}+\frac{\sigma_{v}^{2}}{2 \sigma_{u}^{2}}\right\}}{\sigma_{u} f_{\varepsilon}(\varepsilon)} \int_{0}^{\infty} g(w) \exp \left\{\frac{-w}{\sigma_{u}}\right\} f_{G}(w ; k, \theta) \frac{1}{\sigma_{u}} \Phi\left(-\frac{\varepsilon-w}{\sigma_{v}}-\frac{\sigma_{v}}{\sigma_{u}}\right) d w . \tag{3.86}
\end{equation*}
$$

Note that in the right hand side, $w$ is just the dummy variable of integration. Note also that when $g(w)=e^{w}$, and given that the integral in the numerator does not include a normal density, for it to converge we require

$$
\begin{equation*}
1-1 / \delta<0 \Rightarrow \delta<1 \Rightarrow \frac{\sigma_{u} \theta}{\sigma_{u}+\theta}<1 \Rightarrow \sigma_{u} \theta<\sigma_{u}+\theta \tag{3.87}
\end{equation*}
$$

so that the exponential term in the integrand has a negative power. This is guaranteed if at least one of the two parameters is smaller than unity (which is usually the case). If both are greater than unity, then the required inequality will hold if $\sigma_{u}<\theta /(\theta-1)$.

For the negative error component we obtained

$$
\begin{equation*}
E(u \mid \varepsilon)=\frac{\left(\sigma_{v} / \sigma_{u}\right)}{f_{\varepsilon}(\varepsilon)} \int_{0}^{\infty} f_{G}(u ; k, \theta) \phi\left(\frac{\varepsilon-u}{\sigma_{v}}\right)\left[1-\frac{\psi \Phi(-\psi)}{\phi(\psi)}\right] d u, \quad \psi=\frac{\varepsilon-u}{\sigma_{v}}+\frac{\sigma_{v}}{\sigma_{u}}, \tag{3.88}
\end{equation*}
$$

and

$$
\begin{align*}
E(\exp \{ \pm u\} \mid \varepsilon)= & \frac{\exp \left\{\frac{\sigma_{v}^{2}}{2}\left(\sigma_{u}^{-1} \mp 1\right)^{2}+\left(\sigma_{u}^{-1} \mp 1\right) \varepsilon\right\}}{\sigma_{u} f_{\varepsilon}(\varepsilon)} \times  \tag{3.89}\\
& \times \int_{0}^{\infty} \exp \left\{-\left(\sigma_{u}^{-1} \mp 1\right) w\right\} f_{G}(w ; k, \theta) \Phi\left(-\left(\frac{\varepsilon-w}{\sigma_{v}}+\frac{\sigma_{v}}{\sigma_{u}} \mp \sigma_{v}\right)\right) d w .
\end{align*}
$$

For the case of $\exp \{u\}$, the same condition as before related to $\delta$ is required for the convergence of the integral in the numerator.

Finally the net effect $E\left(e^{w-u} \mid \varepsilon\right)$ is

$$
\begin{align*}
E\left(e^{w-u} \mid \varepsilon\right)=\frac{\sigma_{u}^{k-1}}{f_{\varepsilon}(\varepsilon)\left(\sigma_{u}+\theta\right)^{k}}[ & \exp \left\{\frac{\sigma_{v}^{2}}{2}\left(1+\sigma_{u}^{-1}\right)^{2}+\left(1+\sigma_{u}^{-1}\right) \varepsilon\right\}  \tag{3.90}\\
& \left.-\frac{1}{\sigma_{v}} \int_{0}^{\infty} \exp \left\{\left(1+\sigma_{u}^{-1}\right) z\right\} \phi\left(\frac{\varepsilon-z}{\sigma_{v}}\right) F_{G}(z ; k, \delta) d z\right] .
\end{align*}
$$

## IV.1.4. Quadrature.

## IV.1.4.1. Gauss-Laguerre quadrature.

The $(0, \infty)$ integrating limits in the above expressions and the existence of exponentials in the integrands suggest the use of Gauss-Laguerre quadrature, which approximates integrals of the form

$$
\begin{equation*}
\int_{0}^{\infty} e^{-w} p(w) d x \approx \sum_{j=1}^{m} d_{j} p\left(w_{j}\right) . \tag{3.91}
\end{equation*}
$$

Depending on the specific measure we want to obtain in each case, we will have to bring the obtained expressions into the above integrand form in order to determine the $p\left(x_{j}\right)$ function to be used in each case (by adding and subtracting the exponential discount term where needed) and also, obtain pairs $\left(d_{j}, w_{j}\right)$ of weights $d_{j}$ and nodes (values in the support) $w_{j}, j=1, \ldots, m$. These pairs are not set arbitrarily, and the most widely used method of obtaining them is the one developed by Golub and Welch (1969). The number of nodes to be used (the value of $m$ ) is not standard in the literature, and one can find values ranging from $m=15$ to $m=40$. This is what makes Gauss-Laguerre quadrature a more efficient method compared to the more primitive Newton-Cotes approach like the trapezoid rule, that requires a fine grid and therefore many more evaluation points in order to achieve an acceptable accuracy. Still, the trapezoid rule, being simpler, is more robust.

Using the following expressions, and plugging in the estimated parameters, and for every $i$ the corresponding corrected residual $\hat{\varepsilon}_{i, C O L S}$, we can compute the two integrals involved to obtain the estimated conditional expected value, repeating for each $i$.

The computable expressions are:

$$
\begin{equation*}
\hat{E}\left(w_{i} \mid \varepsilon_{i}\right)=\frac{\exp \left\{\frac{\hat{\varepsilon}_{i, C O L S}}{\hat{\sigma}_{u}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{u}^{2}}\right\} \sum_{j=1}^{m} d_{j} w_{j} e^{w_{j}} e^{-w_{j} / \hat{\sigma}_{u}} f_{G}\left(w_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left(\frac{w_{j}}{\hat{\sigma}_{v}}-\left(\frac{\hat{\varepsilon}_{i, \text { COLS }}}{\hat{\sigma}_{v}}+\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{u}}\right)\right)}{\hat{\sigma}_{u} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { COLS }}\right)}, \tag{3.92}
\end{equation*}
$$

$$
\begin{equation*}
\hat{E}\left(e^{w} \mid \varepsilon_{i}\right)=\frac{\exp \left\{\frac{\hat{\varepsilon}_{i, C O L S}}{\hat{\sigma}_{u}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{u}^{2}}\right\} \sum_{j=1}^{m} d_{j} e^{2 w_{j}} e^{-w_{j} / \hat{\sigma}_{u}} f_{G}\left(w_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left(\frac{w_{j}}{\hat{\sigma}_{v}}-\left(\frac{\hat{\varepsilon}_{i, \text { COLS }}}{\hat{\sigma}_{v}}+\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{u}}\right)\right)}{\hat{\sigma}_{u} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { COLS }}\right)}, \tag{3.93}
\end{equation*}
$$

$$
\begin{equation*}
\hat{E}\left(e^{-w} \mid \varepsilon_{i}\right)=\frac{\exp \left\{\frac{\hat{\varepsilon}_{i, \text { CoLS }}}{\hat{\sigma}_{u}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{u}^{2}}\right\} \sum_{j=1}^{m} d_{j} e^{-w_{j} / \hat{\sigma}_{u}} f_{G}\left(w_{j} ; \hat{k}, \hat{\delta}\right) \Phi\left(\frac{w_{j}}{\hat{\sigma}_{v}}-\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}}{\hat{\sigma}_{v}}+\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{u}}\right)\right)}{\hat{\sigma}_{u} f_{\varepsilon}\left(\hat{\varepsilon}_{i, C o L S}\right)} \tag{3.94}
\end{equation*}
$$

$$
\begin{aligned}
& \hat{E}\left(u_{i} \mid \varepsilon_{i}\right)=\frac{\left(\hat{\sigma}_{v} / \hat{\sigma}_{u}\right)}{f_{\varepsilon}\left(\hat{\varepsilon}_{i, C O L S}\right)}\left\{\sum_{j=1}^{m} d_{j} e^{u_{j}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \phi\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}-u_{j}}{\hat{\sigma}_{v}}\right)\left[1-\frac{\hat{\psi}_{j i} \Phi\left(-\hat{\psi}_{j i}\right)}{\phi\left(\hat{\psi}_{j i}\right)}\right]\right\} \\
& \hat{\psi}_{j i}=\frac{\hat{\varepsilon}_{i, \text { CoLS }}-u_{j}}{\hat{\sigma}_{v}}+\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{u}}, \\
& \hat{E}\left(e^{u_{i}} \mid \varepsilon_{i}\right)=\exp \left\{\frac{\hat{\sigma}_{v}^{2}}{2}\left(\hat{\sigma}_{u}^{-1}-1\right)^{2}+\left(\hat{\sigma}_{u}^{-1}-1\right) \hat{\varepsilon}_{i, \text { CoLs }}\right\}
\end{aligned}
$$

$$
\begin{equation*}
\times \frac{\sum_{j=1}^{m} d_{j} e^{2 u_{j}} e^{-u_{j} / \sigma_{u}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left[-\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}-u_{j}}{\hat{\sigma}_{v}}+\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{u}}-\hat{\sigma}_{v}\right)\right]}{\hat{\sigma}_{u} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)}, \tag{3.96}
\end{equation*}
$$

$$
\begin{align*}
& \hat{E}\left(e^{-u_{i}} \mid \varepsilon_{i}\right)=\exp \left\{\frac{\hat{\sigma}_{v}^{2}}{2}\left(\hat{\sigma}_{u}^{-1}+1\right)^{2}+\left(\hat{\sigma}_{u}^{-1}+1\right) \hat{\varepsilon}_{i, \text { CoLs }}\right\} \\
& \times \frac{\sum_{j=1}^{m} d_{j} e^{-u_{j} / \sigma_{u}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left[-\left(\frac{\hat{\varepsilon}_{i, \text { CoLs }}-u_{j}}{\hat{\sigma}_{v}}+\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{u}}+\hat{\sigma}_{v}\right)\right]}{\hat{\sigma}_{u} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)},  \tag{3.97}\\
& \begin{array}{c}
\hat{E}\left(e^{w_{i}-u_{i}} \mid \varepsilon_{i}\right)=\frac{\hat{\sigma}_{u}^{\hat{k}-1}}{f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)\left(\hat{\sigma}_{u}+\hat{\theta}\right)^{\hat{k}}}\left[\exp \left\{\frac{\hat{\sigma}_{v}^{2}}{2}\left(1+\hat{\sigma}_{u}^{-1}\right)^{2}+\left(1+\hat{\sigma}_{u}^{-1}\right) \hat{\varepsilon}_{i, \text { CoLs }}\right\}\right. \\
\left.-\sum_{j=1}^{m} d_{j} e^{2 w_{j}} e^{w_{j} / \hat{\sigma}_{u}} \frac{1}{\hat{\sigma}_{v}} \phi\left(\frac{\hat{\varepsilon}_{i, \text { CoLs }}-w_{j}}{\hat{\sigma}_{v}}\right) F_{G}\left(w_{j} ; \hat{k}, \hat{\delta}\right)\right],
\end{array} \tag{3.98}
\end{align*}
$$

where

$$
\begin{align*}
f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLs }}\right)= & \frac{\hat{\sigma}_{u}^{\hat{k}-1}}{\left(\hat{\sigma}_{u}+\hat{\theta}\right)^{\hat{k}}}\left[\exp \left\{\frac{\hat{\varepsilon}_{i, \text { CoLS }}}{\hat{\sigma}_{u}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{u}^{2}}\right\}\right.  \tag{3.99}\\
& \left.-\sum_{j=1}^{m} d_{j} e^{w_{j}} e^{w_{j} / \hat{\sigma}_{u}} \frac{1}{\hat{\sigma}_{v}} \phi\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}-w_{j}}{\hat{\sigma}_{v}}\right) F_{G}\left(w_{j} ; \hat{k}, \hat{\delta}\right)\right] .
\end{align*}
$$

## IV.1.4.1. Newton-Cotes quadrature (trapezoid rule).

Experimentation with various data sets revealed problems with Gauss-Laguerre quadrature. In many cases, we obtained inadmissible results (like for example negative values for a density). So, although more resource-intensive and primitive, a Newton-Cotes approach is always an alternative, specifically in its "trapezoid rule" variant.

For this case we note the following: for all integrals to be evaluated, the integrand takes the value zero when the variable of integration equals zero and also, it tends to zero as the variable of integration goes to infinity.

So to apply the trapezoid rule, we suggest that the first point of evaluation is zero, while the last, maximum value for the integrating variable, should be such so as to result in
the integrand taking a value sufficiently close to zero (say, at the order of seven or nine negative powers of 10 ).

Regarding the step-length $h$ between evaluation points, depending also on the whole evaluation interval as determined in the previous paragraph, it could range from $h=0.01$ to $h=0.05$. Then, we approximate the integrals by

$$
\begin{equation*}
\int_{0}^{\infty} f(x) d x \approx h \sum_{j=1}^{m} f\left(x_{j}\right), \quad x_{1}=0, \quad x_{m}: f\left(x_{m}\right) \approx 0 . \tag{3.100}
\end{equation*}
$$

## IV.2. The Exponential-Gamma 2TSF specification.

In this section we present the alternative semi-Gamma specification, where it is the negative one-sided error component that follows a Gamma distribution. Specifically, we have here

$$
v \sim N\left(0, \sigma_{v}^{2}\right), w \sim \operatorname{Exp}\left(\sigma_{w}\right), \quad u \sim \operatorname{Gamma}(k, \theta) .
$$

As before we use the shape-scale parametrization for the Gamma distribution,

$$
\begin{equation*}
f_{u}(u)=\frac{1}{\Gamma(k) \theta^{k}} u^{k-1} \exp \{-u / \theta\} . \tag{3.101}
\end{equation*}
$$

The mean and variance of the composite error term are

$$
\begin{equation*}
E(\varepsilon)=\sigma_{w}-k \theta, \quad \operatorname{Var}(\varepsilon) \equiv \sigma_{\varepsilon}^{2}=\sigma_{v}^{2}+\sigma_{w}^{2}+k \theta^{2} \tag{3.102}
\end{equation*}
$$

## IV.2.1. The Composite error density.

We obtained the composite error density starting with the difference $z=w-u$ for which we got

$$
f_{z}(z)= \begin{cases}\frac{\sigma_{w}^{k-1}}{\left(\sigma_{w}+\theta\right)^{k}} \exp \left\{-z / \sigma_{w}\right\}\left[1-F_{G}(-z ; k, \delta)\right] & z \leq 0  \tag{3.103}\\ \frac{\sigma_{w}^{k-1}}{\left(\sigma_{w}+\theta\right)^{k}} \exp \left\{-z / \sigma_{w}\right\} & z>0\end{cases}
$$

where, $F_{G}(z ; k, \delta)$ is again a Gamma distribution function, here with shape parameter $k$ and scale parameter $\delta \equiv \sigma_{w} \theta /\left(\sigma_{w}+\theta\right)$. The density of the composite error term is

$$
\begin{equation*}
f_{\varepsilon}(\varepsilon)=\frac{\sigma_{w}^{k-1}}{\left(\sigma_{w}+\theta\right)^{k}}\left[\exp \left\{-\frac{\varepsilon}{\sigma_{w}}+\frac{\sigma_{v}^{2}}{2 \sigma_{w}^{2}}\right\}-\int_{0}^{\infty} e^{z / \sigma_{w}} \frac{1}{\sigma_{v}} \phi\left(\frac{\varepsilon+z}{\sigma_{v}}\right) F_{G}(z ; k, \delta) d z\right] . \tag{3.104}
\end{equation*}
$$

## Skewness.

We obtain analogous results with the first variant of the semi-Gamma specification. We have here

$$
\operatorname{sign}\left\{\gamma_{1}(\varepsilon)\right\}=\operatorname{sign}\left\{\kappa_{3}(\varepsilon)\right\}=\operatorname{sign}\left\{\sigma_{w}^{3}-k \theta^{3}\right\},
$$

and as before this sign will not necessarily be the same as the sign of $E(w)-E(u)=\sigma_{w}-k \theta$.
If $k<1$, it may be the case that $k \theta<\sigma_{w}<k^{1 / 3} \theta$ implying that $E(w)-E(u)>0$ but $\operatorname{sign}\left\{\gamma_{1}\right\}<0$.

If $k>1$, we may have $k^{1 / 3} \theta<\sigma_{w}<k \theta$ implying that $E(w)-E(u)<0$ but $\operatorname{sign}\left\{\gamma_{1}\right\}>0$.

## IV.2.2. Moment Equations for the 2TSF Exponential-Gamma specification.

Applying our specific distributional assumptions, the non-linear system of equations that the MM estimator must solve becomes

$$
\begin{cases}\hat{\kappa}_{2}(\varepsilon) & -\left(\sigma_{v}^{2}+\sigma_{w}^{2}+k \theta^{2}\right)=0  \tag{3.105}\\ (1 / 2) \hat{\kappa}_{3}(\varepsilon) & -\left(\sigma_{w}^{3}-k \theta^{3}\right)=0 \\ (1 / 6) \hat{\kappa}_{4}(\varepsilon) & -\left(\sigma_{w}^{4}+k \theta^{4}\right)=0 \\ (1 / 24) \hat{\kappa}_{5}(\varepsilon) & -\left(\sigma_{w}^{5}-k \theta^{5}\right)=0\end{cases}
$$

Setting as before $a_{3}=(1 / 2) \hat{\kappa}_{3}(\varepsilon), a_{4}=(1 / 6) \hat{\kappa}_{4}(\varepsilon), a_{5}=(1 / 24) \hat{\kappa}_{5}(\varepsilon)$ we focus on the $3 \times 3$ subsystem

$$
\left\{\begin{array}{l}
a_{3}-\sigma_{w}^{3}+k \theta^{3}=0  \tag{3.106}\\
a_{4}-\sigma_{w}^{4}-k \theta^{4}=0 \\
a_{5}-\sigma_{w}^{5}+k \theta^{5}=0
\end{array}\right.
$$

Manipulating,

$$
\left\{\begin{array} { l } 
{ k \theta ^ { 3 } = \sigma _ { w } ^ { 3 } - a _ { 3 } } \\
{ k \theta ^ { 3 } = \frac { a _ { 4 } - \sigma _ { w } ^ { 4 } } { \theta } } \\
{ k \theta ^ { 3 } = \frac { \sigma _ { w } ^ { 5 } - a _ { 5 } } { \theta ^ { 2 } } }
\end{array} \Rightarrow \left\{\begin{array}{l}
\theta=\frac{a_{4}-\sigma_{w}^{4}}{\sigma_{w}^{3}-a_{3}} \\
\theta=\frac{\sigma_{w}^{5}-a_{5}}{a_{4}-\sigma_{w}^{4}}
\end{array} \Rightarrow\left(a_{4}-\sigma_{w}^{4}\right)^{2}=\left(\sigma_{w}^{3}-a_{3}\right)\left(\sigma_{w}^{5}-a_{5}\right)\right.\right.
$$

which leads to an equation in only one unknown

$$
\begin{align*}
& a_{4}^{2}-2 a_{4} \sigma_{w}^{4}+\sigma_{w}^{8}=\sigma_{w}^{8}-a_{5} \sigma_{w}^{3}-a_{3} \sigma_{w}^{5}+a_{3} a_{5} \Rightarrow a_{3} \sigma_{w}^{5}-2 a_{4} \sigma_{w}^{4}+a_{5} \sigma_{w}^{3}+a_{4}^{2}-a_{3} a_{5}=0 \\
& \Rightarrow \sigma_{w}^{3}\left(a_{3} \sigma_{w}^{2}-2 a_{4} \sigma_{w}+a_{5}\right)+\left(a_{4}^{2}-a_{3} a_{5}\right)=0 \tag{3.107}
\end{align*}
$$

As in the other variant of the semi-Gamma specification, after running OLS estimation and obtaining the kapa-statistics, we solve this equation and check whether it leads to admissible values for all four unknown parameters, and if it does, these will be used as the starting values for the MM estimator.

Finally, we show in the Technical Appendix that the determinant of the Jacobian of the full system is

$$
\begin{equation*}
|\mathbf{J}|=-2 \sigma_{v} \sigma_{w}^{2} \theta^{6} k\left[3 \theta^{2}+8 \theta \sigma_{w}+5 \sigma_{w}^{2}\right] \tag{3.108}
\end{equation*}
$$

which can be zero only if one of the parameters is zero.

## IV.2.3. Individual JLMS measures.

We obtained the following expressions:

$$
\begin{align*}
& E(w \mid \varepsilon)=\frac{\left(\sigma_{v} / \sigma_{w}\right)}{f_{\varepsilon}(\varepsilon)} \int_{0}^{\infty} f_{G}(u ; k, \theta) \phi\left(\frac{\varepsilon+u}{\sigma_{v}}\right)\left[1-\frac{\psi \Phi(\psi)}{\phi(\psi)}\right] d u, \quad \psi=\frac{\sigma_{v}}{\sigma_{w}}-\frac{(\varepsilon+u)}{\sigma_{v}}  \tag{3.109}\\
& E(\exp \{ \pm w\} \mid \varepsilon)=  \tag{3.110}\\
& \quad \begin{aligned}
& \sigma_{0} f_{\varepsilon}(\varepsilon) \\
& \times \int_{0}^{\infty} \exp \left\{-\left(\sigma_{w}^{-1} \mp 1\right) u\right\} f_{G}^{2}(u ; k, \theta) \Phi\left(-\left(\frac{\sigma_{v}}{\sigma_{w}} \mp \sigma_{v}^{-1}-\frac{(\varepsilon+u)}{\sigma_{v}}\right)\right) d u \\
& E(g(u) \mid \varepsilon)= \frac{\exp \left\{-\frac{\varepsilon}{\sigma_{w}}+\frac{\sigma_{v}^{2}}{2 \sigma_{w}^{2}}\right\}}{\sigma_{w} f_{\varepsilon}(\varepsilon)} \int_{0}^{\infty} g(u) \exp \left\{\frac{-u}{\sigma_{w}}\right\} f_{G}(u ; k, \theta) \Phi\left(\frac{(\varepsilon+u)}{\sigma_{v}}-\frac{\sigma_{v}}{\sigma_{w}}\right) d u
\end{aligned}
\end{align*}
$$

where $f_{G}(u ; k, \theta)$ is a Gamma density and $g(u)=u, \exp \{ \pm u\}$. For the net effect in a semi$\log$ or log-log regression we have

$$
\begin{align*}
E\left(e^{w-u} \mid \varepsilon\right)=\frac{1}{f_{\varepsilon}(\varepsilon)} \frac{\sigma_{w}^{k-1}}{\left(\sigma_{w}+\theta\right)^{k}} & {\left[\exp \left\{\frac{\sigma_{v}^{2}}{2}\left(1-\sigma_{w}^{-1}\right)^{2}+\left(1-\sigma_{w}^{-1}\right) \varepsilon\right\}\right.}  \tag{3.112}\\
& \left.-\int_{0}^{\infty} \exp \left\{-\left(1-\sigma_{w}^{-1}\right) z\right\} \frac{1}{\sigma_{v}} \phi\left(\frac{\varepsilon+z}{\sigma_{v}}\right) F_{G}(z ; k, \delta) d z\right] .
\end{align*}
$$

## IV.2.4. Gauss-Laguerre quadrature.

The computable formulas for the above expressions are

$$
\begin{gather*}
\hat{E}\left(w_{i} \mid \varepsilon_{i}\right)=\frac{\left(\hat{\sigma}_{v} / \hat{\sigma}_{w}\right)}{f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)}\left\{\sum_{j=1}^{m} d_{j} e^{u_{j}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \phi\left(\frac{\hat{\varepsilon}_{i, C o L S}+u_{j}}{\hat{\sigma}_{v}}\right)\left[1-\frac{\hat{\psi}_{j i} \Phi\left(\hat{\psi}_{j i}\right)}{\phi\left(\hat{\psi}_{j i}\right)}\right]\right\}  \tag{3.113}\\
\hat{\psi}_{j i}=\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{w}}-\frac{\hat{\varepsilon}_{i, \text { CoLS }}+u_{j}}{\hat{\sigma}_{v}}+\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{w}},
\end{gather*}
$$

$$
\begin{align*}
\hat{E}\left(e^{w_{i}} \mid \varepsilon_{i}\right)= & \exp \left\{\frac{\hat{\sigma}_{v}^{2}}{2}\left(\hat{\sigma}_{w}^{-1}-1\right)^{2}-\left(\hat{\sigma}_{w}^{-1}-1\right) \hat{\varepsilon}_{i, \text { COLS }}\right\} \\
& \times \frac{\sum_{j=1}^{m} d_{j} e^{2 u_{j}} e^{-u_{j} / \sigma_{w}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left[-\left(\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{w}}-\hat{\sigma}_{v}-\frac{\hat{\varepsilon}_{i, \text { CoLS }}+u_{j}}{\hat{\sigma}_{v}}\right)\right]}{\hat{\sigma}_{w} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)}, \tag{3.114}
\end{align*}
$$

$$
\begin{align*}
& \hat{E}\left(e^{-w_{i}} \mid \varepsilon_{i}\right)=\exp \left\{\frac{\hat{\sigma}_{v}^{2}}{2}\left(\hat{\sigma}_{w}^{-1}+1\right)^{2}-\left(\hat{\sigma}_{w}^{-1}+1\right) \hat{\varepsilon}_{i, \text { CoLs }}\right\}  \tag{3.115}\\
& \times \frac{\sum_{j=1}^{m} d_{j} e^{-u_{j} / \sigma_{u}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left[-\left(\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{w}}+\hat{\sigma}_{v}-\frac{\hat{\varepsilon}_{i, \text { CoLS }}+u_{j}}{\hat{\sigma}_{v}}\right)\right]}{\hat{\sigma}_{w} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)}
\end{align*}
$$

$$
\begin{aligned}
& \hat{E}\left(u_{i} \mid \varepsilon_{i}\right)=\frac{\exp \left\{-\frac{\hat{\varepsilon}_{i, \text { CoLS }}}{\hat{\sigma}_{w}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{w}^{2}}\right\}}{\hat{\sigma}_{w} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)} \times \\
& \quad \times \sum_{j=1}^{m} d_{j} u_{j} e^{u_{j}} e^{-u_{j} / \hat{\sigma}_{w}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}+u_{j}}{\hat{\sigma}_{v}}-\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{w}}\right), \\
& \begin{aligned}
& \hat{E}\left(e^{u_{i}} \mid \varepsilon_{i}\right)= \frac{\exp \left\{-\frac{\hat{\varepsilon}_{i, \text { CoLS }}}{\hat{\sigma}_{w}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{w}^{2}}\right\}}{\hat{\sigma}_{w} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { COLS }}\right)} \times \\
& \quad \times \sum_{j=1}^{m} d_{j} e^{2 u_{j}} e^{-u_{j} / \hat{\sigma}_{w}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}+u_{j}}{\hat{\sigma}_{v}}-\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{w}}\right)
\end{aligned}
\end{aligned}
$$

$$
\hat{E}\left(e^{-u_{i}} \mid \varepsilon_{i}\right)=\frac{\exp \left\{-\frac{\hat{\varepsilon}_{i, \text { CoLS }}}{\hat{\sigma}_{w}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{w}^{2}}\right\}}{\hat{\sigma}_{w} f_{\varepsilon}\left(\hat{\varepsilon}_{i, \text { CoLS }}\right)} \times
$$

$$
\times \sum_{j=1}^{m} d_{j} e^{-u_{j} / \hat{\sigma}_{w}} f_{G}\left(u_{j} ; \hat{k}, \hat{\theta}\right) \Phi\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}+u_{j}}{\hat{\sigma}_{v}}-\frac{\hat{\sigma}_{v}}{\hat{\sigma}_{w}}\right),
$$

$$
\begin{equation*}
\hat{E}\left(e^{w_{i}-u_{i}} \mid \varepsilon_{i}\right)=\frac{\hat{\sigma}_{w}^{\hat{k}-1}}{f_{\varepsilon}\left(\hat{\varepsilon}_{i, C O L S}\right)\left(\hat{\sigma}_{w}+\hat{\theta}\right)^{\hat{k}}}\left[\exp \left\{\frac{\hat{\sigma}_{v}^{2}}{2}\left(1-\hat{\sigma}_{w}^{-1}\right)^{2}+\left(1-\hat{\sigma}_{w}^{-1}\right) \hat{\varepsilon}_{i, \text { CoLS }}\right\}\right. \tag{3.119}
\end{equation*}
$$

$$
\left.-\sum_{j=1}^{m} d_{j} e^{w_{j} / \hat{\sigma}_{w}} \frac{1}{\hat{\sigma}_{v}} \phi\left(\frac{\hat{\varepsilon}_{i, C O L S}+w_{j}}{\hat{\sigma}_{v}}\right) F_{G}\left(w_{j} ; \hat{k}, \hat{\delta}\right)\right],
$$

$$
\begin{equation*}
f_{\varepsilon}\left(\hat{\varepsilon}_{i, C O L S}\right)=\frac{\hat{\sigma}_{w}^{\hat{k}-1}}{\left(\hat{\sigma}_{w}+\hat{\theta}\right)^{\hat{k}}}\left[\exp \left\{-\frac{\hat{\varepsilon}_{i, \text { CoLS }}}{\hat{\sigma}_{w}}+\frac{\hat{\sigma}_{v}^{2}}{2 \hat{\sigma}_{w}^{2}}\right\}\right. \tag{3.120}
\end{equation*}
$$

$$
\left.-\sum_{j=1}^{m} d_{j} e^{w_{j}} e^{w_{j} / \hat{\sigma}_{w}} \frac{1}{\hat{\sigma}_{v}} \phi\left(\frac{\hat{\varepsilon}_{i, \text { CoLS }}+w_{j}}{\hat{\sigma}_{v}}\right) F_{G}\left(w_{j} ; \hat{k}, \hat{\delta}\right)\right]
$$

The comments regarding the applicability of Gauss-Laguerre quadrature and the alternative trapezoid rule in the Gamma-Exponential specification, apply here too.

With these we complete the presentation of the two variants of the semi-Gamma 2TSF specification. We turn to the 2nd empirical application of the chapter.

## IV.3. Empirical application \#2: Implementing the semi-Gamma specification with the COLS-MM estimator.

For the second empirical application of the chapter, we used the data set from Koop and Tobias (2004). This data set was used by Tsionas (2012) to estimate a 2TSF model using the Fast-Fourier Transform. He obtained an Exponential-Gamma specification and we will use the exact same regressor specification. The data come from the National Longitudinal Survey of Youth (NLSY), USA. It is a panel, containing information related to labor earnings on 2,178 white males for a total of 17,919 observations. The data refers to the period 19791993 (fifteen years), and all individuals where in the age 16-22 as of the first interview date in 1979, and not less than 16 in any subsequent year ${ }^{15}$.

The specification adopted by Tsionas (2012) pools the data and includes in the regressors a deterministic time trend and its square. The other regressors are education (in years) and its square, "potential experience" and its square, and a time-invariant measure of "ability", plus a constant term, eight regressors in total (squared variables are multiplied by $1 / 2$ ). The dependent variable is $\log$ hourly wage in 1993 -real terms, so this is a semi-log specification. Our first order of business is to run an OLS regression. This resulted in

[^21]Table 5. OLS estimation results.


The OLS residuals exhibit negative skewness and positive excess kurtosis, specifically we have $\hat{\gamma}_{1}\left(\hat{\varepsilon}_{O L S}\right)=-0.46382$ and $\hat{\gamma}_{2}\left(\hat{\varepsilon}_{O L S}\right)-3=1.5552$. Both support the specification of a 2TSF error density.

Next we estimate the cumulants of the OLS residuals in order to obtain starting values for the MM estimator. Due to the large sample size all three available estimators, the unbiased kapa-statistics, Fisher's $k$-statistics, and the sample cumulants, are expected to give virtually identical results. Indeed we obtained

Table 6. Estimated cumulants.

| Cumulants | Unbiased <br> kapa-statistic | Fisher's <br> $k$-statistic | Sample <br> cumulant |
| :---: | ---: | ---: | ---: |
| 2nd | 0.2229 | 0.2228 | 0.2228 |
| 3d | -0.0488 | -0.0488 | -0.0488 |
| 4th | 0.0773 | 0.0772 | 0.0772 |
| 5th | -0.0574 | -0.0573 | -0.0572 |

Using the kapa-statistics we solve the system of equations related to the two variants of the semi-Gamma specification.

For the Gamma-Exponential one, the system solution starts from the roots of

$$
\sigma_{u}^{3}\left(a_{3} \sigma_{u}^{2}+2 a_{4} \sigma_{u}+a_{5}\right)-\left(a_{4}^{2}-a_{3} a_{5}\right)=0
$$

Given the obtained values, there are two real roots of this equation, but the one produces negative values for the other parameters, while the other produces almost zero values.

For the Exponential- Gamma variant the relevant equation is

$$
\sigma_{w}^{3}\left(a_{3} \sigma_{w}^{2}-2 a_{4} \sigma_{w}+a_{5}\right)+\left(a_{4}^{2}-a_{3} a_{5}\right)=0
$$

For this equation we obtained a single root, which gave admissible values for the other parameters: we got (approximate values) $\sigma_{v}=0.2183, \sigma_{w}=0.2252, k=1.5026, \theta=0.2879$ (so we end up with the specification that was supported by the Fast-Fourier-Transform methodology of Tsionas 2012). These values were used as starting values for the MM estimator (exactly identified GMM). We obtained

## Table 7. Method-of-Moments as GMM results.

```
Model: 1-step GMM, using observations 1-17919
\begin{tabular}{|c|c|c|c|c|}
\hline & estimate & std. error & z & p -value \\
\hline \(\sigma_{v}\) & 0.218291 & 0.0298094 & 7.32 & \(2.43 e-013\) \\
\hline \(\sigma_{w}\) & 0.225261 & 0.00773561 & 29.12 & \(2.01 e-186\) \\
\hline \(k\) & 1.50259 & 0.309111 & 4.86 & \(1.17 e-06\) \\
\hline \(\theta\) & 0.287899 & 0.0187648 & 15.34 & \(3.97 e-053\) \\
\hline
\end{tabular}
GMM criterion: Q = 1.09136e-028 (TQ = 1.9556e-024)
```

As expected the final values are virtually identical to the starting values (since we have an exactly identified system), but the goal here is to obtain standard errors for the estimates.

## IV.3.1. Sample-level measures.

The estimated moments related to the composite error are

Table 8. Estimated Moments of error components

|  | Mean | SD |  |
| ---: | ---: | ---: | ---: |
| $w$-variable (Exponential) | 0.2252 | 0.2252 | Mode $=0$ |
| $u$-variable (Gamma) | 0.4326 | 0.3529 | Mode $=0.1447$ |
| $\varepsilon=v+w-u$ | -0.20733 | 0.4721 | $\hat{\gamma}_{1}(\varepsilon)=-0.4641$ |

Note that the negative error component has a mode considerably away from zero. It is for such cases that the whole approach is justified.

Since we use a semi-log specification, the effects on the wage level are given by the expected values of the exponentiated variables, that can be calculated using the moment generating functions of the Exponential and the Gamma distribution. We got

Table 9. Sample-average effects of informational inefficiency.

| Quantity | Interpretation | Exponential- <br> Gamma <br> 2TSF <br> Specification |
| :---: | :---: | :---: |
| $\hat{E}\left(e^{w}\right)$ | Gross mark-up on perfect- <br> information price due to <br> informational deficiencies from <br> the side of the firm (buyer) | 1.291 |
| $\hat{E}\left(e^{-u}\right)$ | Gross mark-down on perfect- <br> information price due to <br> informational deficiencies from <br> the side of the employee (seller) | 0.684 |
| $\hat{E}\left(e^{w}\right) \hat{E}\left(e^{-u}\right)-1$ | Net effect | -0.1175 |

The forces created due to incomplete information are large on both sides of the market. Regarding the net effect, on average actual wage is close to minus $12 \%$ below the completeinformation wage, a clear informational disadvantage burdening the worker side. We note
that qualitatively this is the opposite result to that obtained by Tsionas (2012). There, the shape parameter of the Gamma distribution for the $u$-variable was estimated as lower than unity, which makes the density "even more" Exponential-like, and leads to an average value lower than that for $w .{ }^{16}$

## IV.3.2. Transaction-level measures.

We proceed to estimate by quadrature the corresponding JLMS measures. We found that Gauss-Laguerre quadrature does not work with this sample, so we implemented the trapezoid rule.

First, we adjust the OLS constant term and the residuals:

$$
\begin{aligned}
& \hat{\beta}_{0, C O L S}=\hat{\beta}_{0, O L S}-\hat{E}(\varepsilon)=0.335641-(-0.20733)=0.542971, \\
& \hat{\varepsilon}_{i, C O L S}=\hat{\varepsilon}_{i, O L S}+\hat{E}(\varepsilon)=\hat{\varepsilon}_{i, O L S}-0.20733 .
\end{aligned}
$$

We then estimate by the trapezoid rule and for each $\hat{\varepsilon}_{i, \text { CoLS }}$, first the composite error density $f_{\varepsilon}(\varepsilon)$, and then the transaction-level measures of interest,

$$
E(\exp \{w\} \mid \varepsilon), E(\exp \{-u\} \mid \varepsilon), E(\exp \{w-u\} \mid \varepsilon) .
$$

Table 10. Statistics for the series of individual measures.

| Conditional measure | Sample <br> mean | Sample <br> median | SD | $\min$ | $\max$ |
| ---: | ---: | ---: | ---: | ---: | :--- |
| $E(\exp \{w\} \mid \varepsilon)$ | 1.3162 | 1.2353 | 0.2538 | 1.1847 | 6.9881 |
| $E(\exp \{-u\} \mid \varepsilon)$ | 0.6717 | 0.7215 | 0.1371 | 0.0642 | 0.7967 |
| $E(\exp \{w-u\} \mid \varepsilon)$ | 0.8827 | 0.8632 | 0.3229 | 0.0725 | 5.7142 |
| $E(\exp \{w\} \mid \varepsilon) E(\exp \{-u\} \mid \varepsilon)$ | 0.8976 | 0.8912 | 0.3054 | 0.0761 | 5.5675 |

[^22]The sample means of $E(\exp \{w\} \mid \varepsilon)$ and $E(\exp \{-u\} \mid \varepsilon)$ differ from the sample-level moment we obtained previously by the use of the moment-generating functions, due to the approximate evaluation of the integrals involved. The difference is small: 1.31 instead of 1.29 for $E(\exp \{w\} \mid \varepsilon)$ and 0.67 instead of 0.68 for $E(\exp \{-u\} \mid \varepsilon)$, which validates the numerical integration results.

Regarding the gross combined effect on the level of the variable, we have also calculated and we present the simple product term $E(\exp \{w\} \mid \varepsilon) E(\exp \{-u\} \mid \varepsilon)$, which ignores the fact that the variables $w$ and $u$, although unconditionally independent, become dependent when conditioned on $\varepsilon$.

The statistics for the two series are close. Looking more closely, we find that $E(\exp \{w\} \mid \varepsilon) E(\exp \{-u\} \mid \varepsilon)$ overestimates the gross net effect in $\sim 80 \%$ of the cases, although not by much: the series of the difference

$$
E(\exp \{w\} \mid \varepsilon) E(\exp \{-u\} \mid \varepsilon)-E(\exp \{w-u\} \mid \varepsilon)
$$

has maximum value 0.034 . More important perhaps is whether the observations change rank according to the two series. We find that not one observation changed rank. We also observed a clear relation between ranking and over/under estimation. Lower ranks are always overestimated by the simple product, and it is the $20 \%$ of observations that occupy the higher ranks that are underestimated (and the higher the rank, the more the underestimation).

In any case, to the degree that $E(\exp \{w-u\} \mid \varepsilon)$ poses no special problems in calculation, it should be preferred, since the rank-invariance result may be sample-specific, and certainly the actual magnitude of the estimated effect may be also of interest.

With this we conclude the second empirical application of this chapter, and we proceed to the last theoretical 2TSF specification.

## V. The Generalized Exponential 2TSF specification.

In section IV we presented the semi-Gamma specification in order to have a model where at least one of the one-sided error terms has its mode away from zero. We used the Gamma distribution because it nests also the Exponential case. But it came at a cost: the resulting density of the composite error term is not closed-form. This required a different estimation method than maximum likelihood, and also, numerical estimations to obtain individual measures of the one-sided terms. And what if we studied a real-world phenomenon where for both the one-sided error components we could make the case that they have their mode away from zero? This would complicate even further estimation and computation of measures. More-over, Ritter and Simar (1997) have advanced the argument that the Gamma distribution may be a risky choice, because its shape parameter (the one that structurally provides its shape flexibility) is weakly identifiable, even for large samples.

Motivated by all these, in this section we develop a 2TSF composite error specification using a specific incarnation of the Generalized Exponential distribution of Gupta and Kundu (1999). In this 2TSF specification both error components have their mode away from zero, while densities have closed-form expressions, so both maximum likelihood and nonnumerical calculation is feasible. The price to pay? The strictly positive mode is an inherent property and it does not nest the Exponential case (so in a sense "Generalized Exponential" is a misnomer). This is a specification that has to be supported by economic and behavioral arguments -and once it does, this statistical inflexibility does not loom large over the model.

## V.1. The Generalized Exponential distribution.

We will assume that the one-sided components in $\varepsilon=v+w-u$ follow a distribution with the following density, say for $w$,

$$
\begin{equation*}
f_{w}(w)=\frac{2}{\theta_{w}} \exp \left\{-w / \theta_{w}\right\}\left(1-\exp \left\{-w / \theta_{w}\right\}\right), \quad \theta_{w}>0, w \geq 0 . \tag{3.121}
\end{equation*}
$$

Note that the density is two times the product of an Exponential density with an Exponential distribution function with the same parameter. There are at least three ways to obtain the above distribution.

First, as a general consequence of the Probability Integral Transform that states that for every continuous random variable $X$ with distribution function $F_{X}(x)$ and density $f_{X}(x)$ we have that $F_{X}(X) \sim U(0,1)$. This then implies that

$$
E\left[F_{X}(X)\right]=\int_{S_{X}} f_{X}(x) F_{X}(x) d x=\frac{1}{2} \Rightarrow \int_{S_{X}} 2 f_{X}(x) F_{X}(x) d x=1
$$

So the function $2 f_{X}(x) F_{X}(x)$ is non-negative in the support of $X$ and integrates to unity over it, therefore it is a density. The density in [3.121] takes this general result and applies it to the Exponential distribution.

Second, $2 f_{X}(x) F_{X}(x)$ is the density of the maximum of two i.i.d random variables indeed, since its distribution function is $\left[F_{X}(x)\right]^{2}$. This may have an economic interpretation and motivation in certain cases.

Third, as we already mentioned, it can be seen as a special case of the "Generalized Exponential" distribution introduced by Gupta and Kundu (1999), with shape parameter equal to 2 (their $\alpha$ ), scale parameter equal to $\theta_{w}$ (their $\lambda$ ) and location parameter equal to zero (their $\mu$ ). We will write $w \sim G E\left(2, \theta_{w}, 0\right)$.

We have the following results (see Gupta \& Kundu 1999 for the general expressions)

$$
\begin{equation*}
E(w)=\frac{3}{2} \theta_{w}, \quad \operatorname{Var}(w)=\frac{5}{4} \theta_{w}^{2}, \quad \text { mode }=\theta_{w} \ln 2, \quad \text { median }=-\theta_{w} \ln (1-1 / \sqrt{2}) . \tag{3.122}
\end{equation*}
$$

Note that the mode of this distribution is equal to the median of an Exponential distribution with the same scale parameter. The distribution has positive skewness for all values of its parameter.

## V.2. Distribution and density functions.

$$
\text { For } \quad \varepsilon=v+w-u \quad \text { with } \quad v \sim N\left(0, \sigma_{v}^{2}\right), w \sim G E\left(2, \theta_{w}, 0\right), u \sim G E\left(2, \theta_{u}, 0\right) \text {, jointly }
$$ independent, we have the density and distribution functions

$$
\begin{align*}
f_{\varepsilon}(\varepsilon)=\frac{2}{\theta_{w}+\theta_{u}}[ & \frac{2 \theta_{u} \exp \left\{a_{u}\right\} \Phi\left(b_{u}\right)}{\theta_{w}+2 \theta_{u}}-\frac{\theta_{u} \exp \left\{2 a_{u}+\left(\sigma_{v} / \theta_{u}\right)^{2}\right\} \Phi\left(b_{u}-\sigma_{v} / \theta_{u}\right)}{2 \theta_{w}+\theta_{u}}  \tag{3.123}\\
& \left.+\frac{2 \theta_{w} \exp \left\{a_{w}\right\} \Phi\left(b_{w}\right)}{2 \theta_{w}+\theta_{u}}-\frac{\theta_{w} \exp \left\{2 a_{w}+\left(\sigma_{v} / \theta_{w}\right)^{2}\right\} \Phi\left(b_{w}-\sigma_{v} / \theta_{w}\right)}{\theta_{w}+2 \theta_{u}}\right],
\end{align*}
$$

and

$$
\begin{align*}
& F_{\varepsilon}(\varepsilon)=\frac{2}{\theta_{w}+\theta_{u}}\left[\frac{2\left(\theta_{w}+\theta_{u}\right)^{3}+\theta_{w} \theta_{u}\left(\theta_{w}+\theta_{u}\right)}{2\left(\theta_{w}+2 \theta_{u}\right)\left(2 \theta_{w}+\theta_{u}\right)} \Phi\left(\frac{\varepsilon}{\sigma_{v}}\right)\right. \\
& +\frac{2 \theta_{u}^{2}}{\theta_{w}+2 \theta_{u}} \exp \left\{a_{u}\right\} \Phi\left(b_{u}\right)-\frac{\theta_{u}^{2}}{2\left(2 \theta_{w}+\theta_{u}\right)} \exp \left\{2 a_{u}+\frac{\sigma_{v}^{2}}{\theta_{u}^{2}}\right\} \Phi\left(b_{u}-\frac{\sigma_{v}}{\theta_{u}}\right)  \tag{3.124}\\
& \left.-\frac{2 \theta_{w}^{2}}{2 \theta_{w}+\theta_{u}} \exp \left\{a_{w}\right\} \Phi\left(b_{w}\right)+\frac{\theta_{w}^{2}}{2\left(\theta_{w}+2 \theta_{u}\right)} \exp \left\{2 a_{w}+\frac{\sigma_{v}^{2}}{\theta_{w}^{2}}\right\} \Phi\left(b_{w}-\frac{\sigma_{v}}{\theta_{w}}\right)\right],
\end{align*}
$$

with

$$
a_{u}=\frac{\varepsilon}{\theta_{u}}+\frac{\sigma_{v}^{2}}{2 \theta_{u}^{2}}, \quad b_{u}=-\left(\frac{\varepsilon}{\sigma_{v}}+\frac{\sigma_{v}}{\theta_{u}}\right), \quad a_{w}=\frac{\sigma_{v}^{2}}{2 \theta_{w}^{2}}-\frac{\varepsilon}{\theta_{w}}, \quad b_{w}=\frac{\varepsilon}{\sigma_{v}}-\frac{\sigma_{v}}{\theta_{w}} .
$$

Compare these with the shorthands in eq. [3.2]. The basic moments of the composite error term are

$$
E(\varepsilon)=\frac{3}{2}\left(\theta_{w}-\theta_{u}\right), \quad \operatorname{Var}(\varepsilon)=\sigma_{v}^{2}+\frac{5}{4}\left(\theta_{w}^{2}+\theta_{u}^{2}\right), \quad \kappa_{3}(\varepsilon)=\frac{27}{8}\left(\theta_{w}^{3}-\theta_{u}^{3}\right) .
$$

We see that the sign of skewness depends on the $\operatorname{sign}\left\{\theta_{w}-\theta_{u}\right\}$.

The distribution of $z=w-u$ has density

$$
f_{z}(z)= \begin{cases}\frac{2 \theta_{u}}{\theta_{w}+\theta_{u}}\left[\frac{2 \exp \left\{z / \theta_{u}\right\}}{\theta_{w}+2 \theta_{u}}-\frac{\exp \left\{2 z / \theta_{u}\right\}}{2 \theta_{w}+\theta_{u}}\right] & z \leq 0  \tag{3.125}\\ \frac{2 \theta_{w}}{\theta_{w}+\theta_{u}}\left[\frac{2 \exp \left\{-z / \theta_{w}\right\}}{2 \theta_{w}+\theta_{u}}-\frac{\exp \left\{-2 z / \theta_{w}\right\}}{\theta_{w}+2 \theta_{u}}\right] & z>0\end{cases}
$$

and distribution function

$$
F_{z}(z)= \begin{cases}\frac{4 \theta_{u}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{z / \theta_{u}\right\}}{\left(\theta_{w}+2 \theta_{u}\right)}-\frac{\theta_{u}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{2 z / \theta_{u}\right\}}{\left(2 \theta_{w}+\theta_{u}\right)} & z \leq 0  \tag{3.126}\\ 1-\frac{4 \theta_{w}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{-z / \theta_{w}\right\}}{2 \theta_{w}+\theta_{u}}+\frac{\theta_{w}^{2}}{\left(\theta_{w}+\theta_{u}\right)} \frac{\exp \left\{-2 z / \theta_{w}\right\}}{\left(\theta_{w}+2 \theta_{u}\right)} & z>0\end{cases}
$$

## V.3. Measures of interest.

In the SF literature, measures of inefficiency at the sample level as well as at the observation level are usually based on conditional expected values. Conditional or not, the expected value is the theoretical analogue of the arithmetic mean. They are not the only ones proposed, but they are those that are predominantly used.

The expected-value measures are well suited for policy issues, since at the very minimum a government agency needs to know what happens "on average" in a market. Moreover, the conditional expected values used for constructing measures at the observation level are minimum squared-error predictors, and so again valid for a more fine-tuned analysis of the market.

But as tools to be used by the firms themselves, especially for short-term decisions, expected values become less useful: from the point of view of an economic decision-making unit, what is "more likely" to happen is more important for the firm's decisions than what happens in the whole market "on average" (although this knowledge too has its value for the firm).

While densities do not measure probabilities, it is also true that the mode of a distribution (the argmax of its density) is the center of the most probable interval of values of a given length, and so serves as a reasonable location measure of "what is more likely to happen", give or take a few. And indeed, the conditional mode has been proposed alongside the conditional expected value in the seminal Jondrow et al. (1982) paper, as an alternative measure for SF models.

We decided to develop the 2TSF Generalized Exponential specification in order to properly model real-world cases where the mode of the one-sided components is expected to be strictly positive; it appears only natural then to develop measures based on the mode for this specification. This will also allow us to bring in the surface an important interpretational issue that essentially highlights the distinction between causality and joint variation.

We also note that the two approaches are not antagonistic: the measurement concepts we are interested in are random variables with a distribution on their own. Providing both the mode and the expected value of this distribution offers a richer picture and a better understanding of how asymmetries and outliers may affect the outcomes. In fact, once we take that road, we might as well provide also the median and other characteristics of these random variables.

On the practical side, the closed-form expressions for the conditional expected values for this specification proved to be almost longer than a page, and it is doubtful whether such complicated formulas are worth the coding and the calculation. As a compromise, we will proceed as follows: We will provide expressions to calculate modes and medians, both at the sample and at the observation level, as well as for specifications where the dependent variable enters either in levels or in logarithmic form. And we will provide the JLMS expected-value measures up to one final integration, leaving it there for the interested user to go on and either perform this final integration, or evaluate the integrals numerically.

We group measures per regression specification, since this is how they are going to be selected.

## V.3.1. Dependent variable enters in levels.

## V.3.1.1. Sample-level measures.

## A. Modes.

The marginal modes for the one-sided error components have been given previously. The mode of their difference is

$$
\begin{equation*}
\operatorname{mode}(w-u)=\operatorname{mode}(z)=\ln \left(\frac{2 \theta_{w}+\theta_{u}}{\theta_{w}+2 \theta_{u}}\right) \cdot \max \left\{\theta_{w}, \theta_{u}\right\} \tag{3.127}
\end{equation*}
$$

One can verify that $\operatorname{mode}(w-u) \neq \operatorname{mode}(w)-\operatorname{mode}(u)$ and in fact that

$$
|\operatorname{mode}(w-u)|<|\operatorname{mode}(w)-\operatorname{mode}(u)|, \text { except if } \theta_{w}=\theta_{u} .
$$

This may create a bit of a challenge to our understanding: the variables are assumed statistically independent and implicitly not-related in a causal sense also. Even in such a case of full disassociation, we have the result that the "most likely" value of their difference is not equal to the difference of their separate most likely values. This result would not surprise a statistician: there are only few cases (the Normal distribution is the easy example that comes to mind), where the mode of the density of a linear combination of random variables is equal to the linear combination of the modes of the marginal densities.

But what interpretation can we give when the equality does not hold? And if, presumably, the more accurate depiction of the situation is given by the mode of the difference, what is the usefulness then of the marginal modes?

The marginal modes reflect the most likely value of each component, if it operated without the presence of the other. This is a useful piece of information. Also, we will see in chapter 6 that it is the marginal modes that remain in various measures of interest expressed as percentages of the dependent variable, when we use a logarithmic specification. But to
measure the most likely net effect, it is the mode of the distribution of their difference that matters, since, even if independent and not causally related, the two error components happen concurrently and so we must consider their variation jointly.

## B. Medians.

The marginal medians have been given already. To obtain the median of the difference $\operatorname{med}(w-u)$, we first determine whether it is located below or above zero, by evaluating

$$
\operatorname{Pr}(z \leq 0)=F_{z}(0)=\frac{4 \theta_{u}^{2}}{\theta_{w}+\theta_{u}} \frac{1}{\left(\theta_{w}+2 \theta_{u}\right)}-\frac{\theta_{u}^{2}}{\theta_{w}+\theta_{u}} \frac{1}{\left(2 \theta_{w}+\theta_{u}\right)} .
$$

If $\operatorname{Pr}(z \leq 0)>1 / 2$, then $\operatorname{med}(z)<0$ while if $\operatorname{Pr}(z \leq 0)<1 / 2$, then $\operatorname{med}(z)>0$. Then, we must numerically solve the relevant one of the following equations,

$$
\left\{\begin{array}{l}
\frac{4 \theta_{u}^{2}}{\theta_{w}+\theta_{u}} \frac{\exp \left\{\operatorname{med}(z) / \theta_{u}\right\}}{\left(\theta_{w}+2 \theta_{u}\right)}-\frac{\theta_{u}^{2}}{\theta_{w}+\theta_{u}} \frac{\exp \left\{2 \operatorname{med}(z) / \theta_{u}\right\}}{\left(2 \theta_{w}+\theta_{u}\right)}-\frac{1}{2}=0 \quad \operatorname{med}(z) \leq 0  \tag{3.128}\\
\frac{4 \theta_{w}^{2}}{\theta_{w}+\theta_{u}} \frac{\exp \left\{-\operatorname{med}(z) / \theta_{w}\right\}}{2 \theta_{w}+\theta_{u}}-\frac{\theta_{w}^{2}}{\theta_{w}+\theta_{u}} \frac{\exp \left\{-2 \operatorname{med}(z) / \theta_{w}\right\}}{\left(\theta_{w}+2 \theta_{u}\right)}-\frac{1}{2}=0 \quad \operatorname{med}(z)>0
\end{array}\right.
$$

## C. Expected values.

Unconditional expected values for the levels of the one-sided error components were given previously, and due to linearity, the expected value of their difference equals the difference of their expected values.

## V.3.1.2. Observation-level measures.

## A. Conditional Modes.

The conditional modes series are not given in closed form. But it is simple to obtain them if we have the conditional density, since it just requires to calculate, for every value of
$\varepsilon$, the value of the density for some reasonably fine grid of points and keep the point that gives the density's maximum. There is no integral evaluation involved.

We have,

1. For the mode $(w \mid \varepsilon)$

$$
\begin{align*}
f_{w \mid \varepsilon}(w \mid \varepsilon)=\frac{f_{w}(w)}{f_{\varepsilon}(\varepsilon)} \frac{2}{\theta_{u}}\left[\exp \left\{-\frac{w}{\theta_{u}}\right\}\right. & \exp \left\{a_{u}\right\} \Phi\left(\frac{w}{\sigma_{v}}+b_{u}\right) \\
& \left.\quad-\exp \left\{-\frac{2 w}{\theta_{u}}\right\} \exp \left\{2 a_{u}+\frac{\sigma_{v}^{2}}{\theta_{u}^{2}}\right\} \Phi\left(\frac{w}{\sigma_{v}}+b_{u}-\frac{\sigma_{v}}{\theta_{u}}\right)\right] . \tag{3.129}
\end{align*}
$$

2. For the $\operatorname{mode}(u \mid \varepsilon)$

$$
\begin{align*}
f_{u \mid \varepsilon}(u \mid \varepsilon)=\frac{f_{u}(u)}{f_{\varepsilon}(\varepsilon)} \frac{2}{\theta_{w}}\left[\exp \left\{-\frac{u}{\theta_{w}}\right\}\right. & \exp \left\{a_{w}\right\} \Phi\left(\frac{u}{\sigma_{v}}+b_{w}\right) \\
& \left.\quad-\exp \left\{-\frac{2 u}{\theta_{w}}\right\} \exp \left\{2 a_{w}+\frac{\sigma_{v}^{2}}{\theta_{w}^{2}}\right\} \Phi\left(\frac{u}{\sigma_{v}}+b_{w}-\frac{\sigma_{v}}{\theta_{w}}\right)\right] . \tag{3.130}
\end{align*}
$$

3. For the $\operatorname{mode}(w-u \mid \varepsilon)=\operatorname{mode}(z \mid \varepsilon)$

$$
f_{z \mid \varepsilon}(z \mid \varepsilon)= \begin{cases}\frac{1}{\sigma_{v} f_{\varepsilon}(\varepsilon)} \phi\left(\frac{\varepsilon-z}{\sigma_{v}}\right) \frac{2 \theta_{u}}{\theta_{w}+\theta_{u}}\left[\frac{2 \exp \left\{z / \theta_{u}\right\}}{\theta_{w}+2 \theta_{u}}-\frac{\exp \left\{2 z / \theta_{u}\right\}}{2 \theta_{w}+\theta_{u}}\right] & z \leq 0  \tag{3.131}\\ \frac{1}{\sigma_{v} f_{\varepsilon}(\varepsilon)} \phi\left(\frac{\varepsilon-z}{\sigma_{v}}\right) \frac{2 \theta_{w}}{\theta_{w}+\theta_{u}}\left[\frac{2 \exp \left\{-z / \theta_{w}\right\}}{2 \theta_{w}+\theta_{u}}-\frac{\exp \left\{-2 z / \theta_{w}\right\}}{\theta_{w}+2 \theta_{u}}\right] & z>0\end{cases}
$$

Due to the existence of branches, to determine the conditional mode here (for each given value of $\varepsilon$ ), we need to calculate both branches over their respective domains, and check which of the two obtained modes is permissible, i.e. which falls in the prescribed domain of each branch (see the Technical Appendix for more details).

## B. Conditional expected values.

Given the conditional densities provided above, we have
$E(w \mid \varepsilon)=\int_{0}^{\infty} w f_{w \mid \varepsilon}(w \mid \varepsilon) d w, \quad E(u \mid \varepsilon)=\int_{0}^{\infty} u f_{u \mid \varepsilon}(u \mid \varepsilon) d u$,
and for the difference $z=w-u, E(z \mid \varepsilon)=\int_{-\infty}^{\infty} z f_{z \mid \varepsilon}(z \mid \varepsilon) d z$.

These integrals can be tediously solved, or be evaluated numerically.

## V.3.2. Dependent variable enters in logarithmic form.

## V.3.2.1. Sample-level measures.

## A. Modes.

Turning to the modes that are pertinent when the dependent variable is in logarithmic form, we have

$$
\begin{equation*}
\operatorname{mode}(\exp \{w\})=\left(\frac{2+\theta_{w}}{1+\theta_{w}}\right)^{\theta_{w}}, \quad \operatorname{mode}(\exp \{-w\})=\left(\frac{1-\theta_{w}}{2-\theta_{w}}\right)^{\theta_{w}} \tag{3.132}
\end{equation*}
$$

and correspondingly for the $u$ variable by simply changing subscripts. For the mode of the net effect mode $(\exp \{w-u\})$ we have

$$
\begin{align*}
& \operatorname{mode}(\exp \{w-u\})=\max \left\{q_{0} \cdot I\left\{q_{0} \leq 1\right\}, q_{1} \cdot I\left\{q_{1}>1\right\}\right\},  \tag{3.133}\\
& q_{0}=\left(\frac{1-\theta_{u}}{2-\theta_{u}} \frac{4 \theta_{w}+2 \theta_{u}}{\theta_{w}+2 \theta_{u}}\right)^{\theta_{u}}, \quad q_{1}=\left(\frac{2+\theta_{w}}{1+\theta_{w}} \frac{2 \theta_{w}+\theta_{u}}{2 \theta_{w}+4 \theta_{u}}\right)^{\theta_{w}} .
\end{align*}
$$

Here $I\{\cdot\}$ is the indicator function. So we need to calculate both expressions to determine the mode. This is due to the fact that the density of $\exp \{w-u\}$ has branches (see the Technical Appendix for details). Here too we observe that

$$
\operatorname{mode}(\exp \{w-u\}) \neq \operatorname{mode}(\exp \{w\}) \cdot \operatorname{mode}(\exp \{-u\})
$$

and the same commentary as before holds.

## B. Medians.

Medians are quantiles and so it holds that $\operatorname{med}(\exp \{w\})=\exp \{\operatorname{med}(w)\}$, and analogously for $u$ and $z=w-u$. So we have these once we calculate the medians of the distributions of $w, u, w-u$.

## C. Expected values.

The unconditional expected values are

$$
\begin{align*}
& E(\exp \{w\})=\frac{2}{\left(1-\theta_{w}\right)\left(2-\theta_{w}\right)}, \quad\left\{t<1 / \theta_{w}\right\} \cup\left\{t>2 / \theta_{w}\right\},  \tag{3.134}\\
& E(\exp \{-w\})=\frac{2}{\left(1+\theta_{w}\right)\left(2+\theta_{w}\right)}, \tag{3.135}
\end{align*}
$$

and analogously for the $u$ variable. Since the variables are assumed independent we also have that

$$
E(\exp \{w-u\})=E(\exp \{w\}) E(\exp \{-u\})
$$

## V.3.2.2. Observation-level measures.

## A. Conditional Modes.

Here we have the following conditional densities:

1. For the $\operatorname{mode}(\exp \{w\} \mid \varepsilon), q=\exp \{w\}$

$$
\begin{align*}
f_{q \mid \varepsilon}(q \mid \varepsilon)=\frac{4}{\theta_{w} \theta_{u}} \frac{\left(q^{-1 / \theta_{w}}-q^{-2 / \theta_{w}}\right)}{f_{\varepsilon}(\varepsilon)}[ & q^{-1 / \theta_{u}-1} \exp \left\{a_{u}\right\} \Phi\left(\frac{\ln q}{\sigma_{v}}+b_{u}\right)  \tag{3.136}\\
& \left.-q^{-2 / \theta_{u}-1} \exp \left\{2 a_{u}+\frac{\sigma_{v}^{2}}{\theta_{u}^{2}}\right\} \Phi\left(\frac{\ln q}{\sigma_{v}}+b_{u}-\frac{\sigma_{v}}{\theta_{u}}\right)\right] .
\end{align*}
$$

2. For the $\operatorname{mode}(\exp \{-w\} \mid \varepsilon), q=\exp \{-w\}$

$$
\begin{align*}
& f_{q \mid \varepsilon}(q \mid \varepsilon)=\frac{4}{\theta_{w} \theta_{u}} \frac{\left(q^{1 / \theta_{w}-1}-q^{2 / \theta_{w}-1}\right)}{f_{\varepsilon}(\varepsilon)}\left[q^{1 / \theta_{u}} \exp \left\{a_{u}\right\} \Phi\left(\frac{-\ln q}{\sigma_{v}}+b_{u}\right)\right. \\
&\left.-q^{2 / \theta_{u}} \exp \left\{2 a_{u}+\frac{\sigma_{v}^{2}}{\theta_{u}^{2}}\right\} \Phi\left(\frac{-\ln q}{\sigma_{v}}+b_{u}-\frac{\sigma_{v}}{\theta_{u}}\right)\right] . \tag{3.137}
\end{align*}
$$

3. For the $\operatorname{mode}(\exp \{-u\} \mid \varepsilon), \quad q=\exp \{-u\}$

$$
\begin{align*}
f_{q \mid \varepsilon}(q \mid \varepsilon)= & \frac{1}{f_{\varepsilon}(\varepsilon)} \frac{4}{\theta_{w} \theta_{u}}\left(q^{1 / \theta_{u}-1}-q^{2 / \theta_{u}-1}\right)\left[q^{1 / \theta_{w}} \exp \left\{a_{w}\right\} \Phi\left(\frac{-\ln q}{\sigma_{v}}+b_{w}\right)\right. \\
& \left.-q^{2 / \theta_{w}} \exp \left\{2 a_{w}+\frac{\sigma_{v}^{2}}{\theta_{w}^{2}}\right\} \Phi\left(\frac{-\ln q}{\sigma_{v}}+b_{w}-\frac{\sigma_{v}}{\theta_{w}}\right)\right] . \tag{3.138}
\end{align*}
$$

4. For the $\operatorname{mode}(\exp \{w-u\} \mid \varepsilon), q=\exp \{w-u\}=\exp \{z\}$

$$
f_{q \mid \varepsilon}(q \mid \varepsilon)=\left\{\begin{array}{lr}
\frac{2 \theta_{u}}{\sigma_{v}\left(\theta_{w}+\theta_{u}\right)} \frac{\phi\left((\varepsilon-\ln q) / \sigma_{v}\right)}{f_{\varepsilon}(\varepsilon)}\left[\frac{2 q^{1 / \theta_{u}-1}}{\theta_{w}+2 \theta_{u}}-\frac{q^{2 / \theta_{u}-1}}{2 \theta_{w}+\theta_{u}}\right] & 0<q \leq 1  \tag{3.139}\\
\frac{2 \theta_{w}}{\sigma_{v}\left(\theta_{w}+\theta_{u}\right)} \frac{\phi\left((\varepsilon-\ln q) / \sigma_{v}\right)}{f_{\varepsilon}(\varepsilon)}\left[\frac{2 q^{-1 / \theta_{w}-1}}{2 \theta_{w}+\theta_{u}}-\frac{q^{-2 / \theta_{w}-1}}{\theta_{w}+2 \theta_{u}}\right] & q>1
\end{array}\right.
$$

## B. Conditional expected values.

As before, we can evaluate numerically the following integrals:

$$
\begin{aligned}
& E(\exp \{ \pm w\} \mid \varepsilon)=\int_{0}^{\infty} \exp \{ \pm w\} f_{w \mid \varepsilon}(w \mid \varepsilon) d w, \\
& E(\exp \{ \pm u\} \mid \varepsilon)=\int_{0}^{\infty} \exp \{ \pm u\} f_{u \mid \varepsilon}(u \mid \varepsilon) d u,
\end{aligned}
$$

and for the difference $z=w-u$

$$
E(\exp \{ \pm z\} \mid \varepsilon)=\int_{-\infty}^{\infty} \exp \{ \pm z\} f_{z \mid \varepsilon}(z \mid \varepsilon) d z
$$

## V.3.2.3. Distributional Connections.

By presenting not just expected values but also the mode and the median of the exponentiated variables, we essentially recognized the fact that they are random variables in their own right. It is useful then to present the distributions that they follow. We have the following results:
A. $q=\exp \{w\}, q \in[1, \infty)$

Remember that $w \sim G E\left(2, \theta_{w}, 0\right)$ can be seen as being the maximum of two i.i.d. Exponential random variables. Then the random variable $q=\exp \{w\}$ is the maximum of
two i.i.d Pareto random variables with lowest value $q_{\min }=1$ and shape parameter $\alpha=1 / \theta_{w}$. Analogously for the $u$ variable.
B. $q=1-\exp \{-w\}, q \in[0,1)$

If $w \sim G E\left(2, \theta_{w}, 0\right)$, then $q=1-\exp \{-w\}$ is the maximum of two i.i.d. Kumaraswamy random variables, with parameters $\alpha=1, \beta=1 / \theta_{w} .{ }^{17}$

Analogously for the $u$ variable.
C. $q=\exp \{-u\}, \quad q \in(0,1]$.

If $u \sim G E\left(2, \theta_{u}, 0\right)$, then $q=\exp \{-u\}$ is the minimum of two i.i.d. Beta random variables with parameters $\alpha=1 / \theta_{w}, \beta=1$.

We have already mentioned earlier that the distributions of $w$ and $u$ are each the distribution of the maximum of two i.i.d variables. So in a notional set up where we may have two outcomes, the 2TSF specification "sees" the strongest. This transfers also to the exponentiated variables. Those that affect positively the outcome are distributed as the maximum of two i.i.d variables, while $\exp \{-u\}$ is distributed as the minimum of two i.i.d. variables, and so again, the realization is the one that has the strongest negative impact on the outcome.

We can say then that the 2TSF Generalized Exponential specification models the real world as operating at maximum intensity, something not incompatible with fundamental ideas in economic theory.

[^23]
## Moving Forward.

Our goal in this chapter was to increase the tools available to estimate 2TSF models with different distributional specifications, specifications that may not even possess a closedform density for the composite error term, or that have their mode away from zero. In this way we can accommodate more accurately a wider class of real-world phenomena.

Throughout the chapter we have assumed that there is no dependence between the components of the error term, and also, that the regressors of the model are exogenous.

Since there are many situations where neither of these assumptions is expected to hold, we relax both in the next chapter and we develop tools to implement the two-tier stochastic frontier framework under error intradependence and regressor endogeneity.-

## Chapter 4

## Dependence and Endogeneity

In this chapter we provide tools to estimate 2TSF models where the two one-sided error components are statistically dependent, and/or the regressors of the model are endogenous, namely correlated with the composite error term. In section I we present the Correlated Exponential 2TSF specification, while in Section II we present a Copula model to handle endogeneity without using instruments, paying attention not only to the bivariate Copula case, but also to the multivariate case where certain new issues arise. Since both models will be used in the empirical studies of later chapters, we do not include in this chapter any empirical application.

## I. The Correlated Exponential 2TSF specification ${ }^{1}$

In almost all stochastic frontier models a maintained assumption is that the two or (three) components of the composite error term are jointly independent. There are many situations where this assumption is rather indefensible. As examples where we should anticipate dependence, consider three situations where the 2TSF model has been applied:

Example 1: In the 2TSF structural framework of Polachek and Yoon (1987), we have interpreted incomplete information in the market as "not knowing the full extent of the market", due to search costs (see chapter 2). For both parties it is costly to acquire this information so not all buyers know of the existence of all sellers, and vice versa. Institutional/structural interventions that reduce such costs will lower the degree of information incompleteness for both parties. For example new "match" technologies like websites where both can post their interest in hiring or be hired together with job descriptions and qualifications will reduce the costs of acquiring information. Statistically, this means that the two one-sided error terms will exhibit positive correlation.

Example 2: In the Nash-bargaining model for the labor market, the relative bargaining power of the worker appears in both one-sided error terms, with opposite sign. Since this relative bargaining power is realistically expected to be a random variable, differing from transaction to transaction, it follows that the two terms are expected to be negatively correlated.

[^24]Example 3: Ferona and Tsionas (2012) estimated underbidding and overbidding behavior in auctions using a 2TSF model. Systematic underbidding and overbidding may be rationalized as "statistical biases" coming from different past experience of the agents. Mathematically both variables representing the underbidding and the overbidding behavior are modeled as non-negative. So "excessive underbidding" means that the one sided error component that enters negatively in the specification takes high values (which are subtracted from some balanced evaluation of the object auctioned).

Now imagine an open auction where underbidding appears "excessive" (offers are much lower than "usual"): this means that the variable representing the underbidding behavior takes higher values. Then systematic overbidders are expected to rationally respond by lowering their overbidding (they don't need to go into higher costs). Mathematically this means that the variable representing the overbidding behavior takes lower values. Here we have negative correlation.

Assume now the opposite situation where overbidding appears "excessive": the variable representing the overbidding behavior takes "higher than usual" values. Systematic underbidders will be forced to increase their offers, which mathematically means that the variable representing the underbidding behavior will take lower values. So in this situation too we expect negative correlation, and so we unambiguously expect negative correlation between the two one-sided error components.

Smith (2008) has developed a single-tier SF model with dependence between the symmetric disturbance term and the inefficiency term, and he has shown that ignoring dependence while it exists may visibly impact the accuracy of the estimates. El Mehdi and Hafner (2014) proposed a Copula to capture such dependence in-between the error components in a single-tier SF model.

In this section we present a 2TSF specification that incorporates dependence between the two one-sided error-terms, while maintaining their independence from the zero-mean symmetric component of the composite error term. Such a dependence structure can be defended if we treat the latter as a conditional expectation function (CEF) error. Specifically, in the abstract 2TSF formulation $y=f(\mathbf{x})+v+w-u$, if we assert that $E[y \mid \mathbf{x}, w, u]=f(\mathbf{x})+w-u$, we have $y=E[y \mid \mathbf{x}, w, u]+v$, and $v$ becomes the CEF error, zero-mean and mean-independent from the regressors but also from $w, u$ by construction:

Although mean-independence does not exclude higher-order statistical dependence, it is the most important one as regards the properties of estimators.

We will base our specification in Freund's (1961) "Bivariate Exponential extension".

## I.1. Freund's Bivariate Exponential extension.

Freund (1961), having in mind a specific real-world situation related to failure and survival times of a two-component system, arrived at the following joint density for two non-negative random variables:

$$
f_{w u}(w, u)=\left\{\begin{array}{ll}
a b^{\prime} \exp \left\{-b^{\prime} u-\left(a+b-b^{\prime}\right) w\right\} & 0<w<u  \tag{4.1}\\
a^{\prime} b \exp \left\{-a^{\prime} w-\left(a+b-a^{\prime}\right) u\right\} & 0<u<w
\end{array} \quad a, a^{\prime}, b, b^{\prime}>0\right.
$$

The distribution is known as "Freund's Bivariate Exponential extension" (see Balakrishnan \& Lai 2009 ch. 10.3 for a collection of literature, properties and applications of the distribution).

The corresponding marginal distributions are
$f_{w}(w)= \begin{cases}\frac{\left(a-a^{\prime}\right)(a+b)}{a+b-a^{\prime}} \exp \{-(a+b) w\}+\frac{a^{\prime} b}{a+b-a^{\prime}} \exp \left\{-a^{\prime} w\right\} & a+b-a^{\prime} \neq 0 \\ \left(a+a^{\prime} b w\right) \exp \left\{-a^{\prime} w\right\} & a+b-a^{\prime}=0\end{cases}$
$f_{u}(u)= \begin{cases}\frac{\left(b-b^{\prime}\right)(a+b)}{a+b-b^{\prime}} \exp \{-(a+b) u\}+\frac{a b^{\prime}}{a+b-b^{\prime}} \exp \left\{-b^{\prime} u\right\} & a+b-b^{\prime} \neq 0 \\ \left(b+b^{\prime} a u\right) \exp \left\{-b^{\prime} u\right\} & a+b-b^{\prime}=0\end{cases}$

We will concentrate on the case $a+b-a^{\prime} \neq 0$ and $a+b-b^{\prime} \neq 0$. The reasons are many: first, it appears rather ad hoc to impose one or both of these coefficient restrictions given that we model here unobservable variables. Second, imposing them implies more restrictions: If we
impose both restrictions, we automatically impose also the restriction $a^{\prime}=b^{\prime}$. If we impose only one of the two, we force the restriction $a^{\prime} \neq b^{\prime}$. Also, the presence of the variable outside the exponential in the lower branches of [4.2] and [4.3], make mathematical derivations and formulas further down the line much more complicated. Finally, the way we will use this distribution strips its parameters from the real-world interpretation they have in the situation for which Freund (1961) created the distribution. So any a priori restriction on them does not correspond to any specific real-world case.

Under $a+b-a^{\prime} \neq 0$ and $a+b-b^{\prime} \neq 0$, each marginal is a mixture (convex combination) of two Exponential densities. Regarding the shape of the densities, generally they are "Exponential-like", in that they have their mode at zero and are monotonically declining with convex curvature. But for some parameter values they become "Gamma-like", with a strictly positive mode and a right tail. So the marginals for this bivariate distribution are more flexible than the Exponential or Half-normal distributions that always have their mode at zero. Compared to the semi-Gamma specification presented also in chapter 3, it allows for both the one-sided terms to have modes away from zero.

For reasons that will be discussed in a while we define the parameter

$$
\begin{equation*}
m \equiv \frac{a}{a+b} . \tag{4.4}
\end{equation*}
$$

The joint moment generating function is ${ }^{2}$

$$
\begin{equation*}
M_{w, u}(s, t)=E(\exp \{s w+t u\})=(a+b-s-t)^{-1}\left[\frac{a^{\prime} b}{a^{\prime}-s}+\frac{a b^{\prime}}{b^{\prime}-t}\right] . \tag{4.5}
\end{equation*}
$$

Using this we obtain the means and variances of the marginals as

[^25]\[

$$
\begin{align*}
& E(w)=\frac{a^{\prime}+b}{a^{\prime}(a+b)}=\frac{1}{a+b}+\frac{1-m}{a^{\prime}} \quad, \quad \operatorname{Var}(w)=\frac{a^{\prime 2}+2 a b+b^{2}}{a^{\prime 2}(a+b)^{2}}=\frac{1}{(a+b)^{2}}+\frac{1-m^{2}}{a^{\prime 2}}  \tag{4.6}\\
& E(u)=\frac{b^{\prime}+a}{b^{\prime}(a+b)}=\frac{1}{a+b}+\frac{m}{b^{\prime}} \quad, \quad \operatorname{Var}(u)=\frac{b^{\prime 2}+2 a b+a^{2}}{b^{\prime 2}(a+b)^{2}}=\frac{1}{(a+b)^{2}}+\frac{m(2-m)}{b^{\prime 2}}
\end{align*}
$$
\]

Note that the variance does not equal the square of the mean. Equivalently, the coefficient of variation is not fixed to unity as with the case of the Exponential distribution, or to $\sqrt{\pi / 2-1} \approx 0.755$ as is the case with the Half-normal distribution. Here, the coefficient of variation can vary and be greater or smaller than unity, depending on the values of the parameters.

## I.1.1. Dependence structure.

Dependence rests on $a \neq a^{\prime}$ or $b \neq b^{\prime}$. If $a=a^{\prime}$ and $b=b^{\prime}$, the two marginals become Exponential, and the variables are independent (with $a^{\prime}, b^{\prime}$ being then the reciprocals of their respective means).

If just one of these equalities holds, then, as Freund (1961) shows, the dependence remains even though one of the marginals becomes Exponential, and even though one of the branches of the joint density is the product of two Exponential densities (but not the product of the densities of the variables involved).

The covariance and Pearson's correlation coefficient are

$$
\begin{align*}
& \operatorname{Cov}(w, u)=\frac{a^{\prime} b^{\prime}-a b}{a^{\prime} b^{\prime}(a+b)^{2}}=\frac{1}{(a+b)^{2}}-\frac{m(1-m)}{a^{\prime} b^{\prime}}  \tag{4.7}\\
& \rho=\frac{a^{\prime} b^{\prime}-a b}{\sqrt{\left(a^{\prime 2}+2 a b+b^{2}\right)\left(b^{\prime 2}+2 a b+a^{2}\right)}} .
\end{align*}
$$

Looking at the expression for covariance in [4.7], we see that the model allows also for the case of only higher-order dependence with zero linear dependence, since the covariance
will be zero if $a^{\prime} b^{\prime}=a b$ which is a weaker condition than $\left(a=a^{\prime}, b=b^{\prime}\right)$ that is required for full independence.

Turning to the correlation coefficient, Moran (1967) has studied the extent to which two non-negative random variables can be negatively correlated, as reflected in the values that the Pearson correlation coefficient can take. Intuitively, while positive correlation can extend all the way to plus unity, negative correlation cannot reach minus unity, because as the one variable tends to infinity the other cannot, by construction, go below zero. So this is not an artificial restriction but reflects the true nature of the relation that two such variables may have. Moran showed that the global maximum negative correlation (i.e. lowest possible, taking the sign into consideration) depends on the marginal distributions irrespective of the joint distribution, and can be calculated by the use of the marginal quantile functions. For example, he found that between two Exponential random variables, the lowest possible value of Pearson's correlation coefficient is $\approx-0.645$.

But the joint distribution is the one that eventually determines the actual exteme correlation values attainable in each case. For Freund's Bivariate Exponential extension, the correlation coefficient lies in $(-1 / 3,1)$. It attains its upper bound when $a^{\prime} \rightarrow \infty, b^{\prime} \rightarrow \infty$. It attains its lower bound when $a^{\prime}, b^{\prime} \rightarrow 0$ and $a=b$ (see Technical Appendix). This correlation interval is for example much wider than what holds for the Gumbel (1960) bivariate Exponential density that uses the Farlie-Morgenstern-Gumbel construction, where the correlation coefficient is restricted to lie in $(-1 / 4,1 / 4)$, or for another proposed joint distribution in the same paper in which the correlation can only be negative.

So apart from providing marginals more flexible than the Exponential and Half-normal specifications, this distribution dominates other "Exponential-based" joint distributions proposed in the literature as regards the extent of correlation that it can accommodate.

Since the marginal distributions are mixtures (convex combinations) of Exponential densities, we call this 2TSF specification "Correlated Exponential".

## I.2. The distribution of the 2TSF Correlated Exponential error term.

Assume that $w, u$ follow jointly Freund's Bivariate Exponential extension, and let $v \sim N\left(0, \sigma_{v}^{2}\right)$, independent of the other two. Then the density of the 2TSF composite error $\operatorname{term} \varepsilon=v+w-u$ is

$$
\begin{align*}
& f_{\varepsilon}(\varepsilon)=\frac{a}{a+b} b^{\prime} \exp \left\{\frac{1}{2} \sigma_{v}^{2} b^{\prime 2}\right\} \exp \left\{b^{\prime} \varepsilon\right\} \Phi\left(-\frac{\varepsilon}{\sigma_{v}}-\sigma_{v} b^{\prime}\right) \\
&  \tag{4.8}\\
& +\frac{b}{a+b} a^{\prime} \exp \left\{\frac{1}{2} \sigma_{v}^{2} a^{\prime 2}\right\} \exp \left\{-a^{\prime} \varepsilon\right\} \Phi\left(\frac{\varepsilon}{\sigma_{v}}-\sigma_{v} a^{\prime}\right)
\end{align*}
$$

If $\left(a^{\prime}, b^{\prime}\right)$ are taken to be the reciprocal of the means of two Exponential random variables, $f_{\varepsilon}(\varepsilon)$ looks very much like the benchmark 2TSF Exponential density (see chapter 3), apart for the terms $a /(a+b)=m$ and $b /(a+b)=1-m$ that multiply the two components of the sum. It is through these terms that statistical dependence is represented. If we set $\left(a^{\prime}=a, b^{\prime}=b\right)$ we obtain the density under independence. So the 2TSF Correlated Exponential composite error density nests the independence case, a desirable property.

## I.2.1. Identification.

The problem we face with the density in [4.8] is that the manner in which the parameters $a, b$ enter the expression, they cannot be separately identified and estimated in a maximum likelihood estimation procedure.

To see this, write the density as $f_{\varepsilon}(\varepsilon)=\frac{a}{a+b} A+\frac{b}{a+b} B$ where $A, B$ are free of $a, b$. Then the log-likelihood is

$$
\begin{aligned}
& \ln L=\sum_{i=1}^{n} \ln \left(\frac{a}{a+b} A_{i}+\frac{b}{a+b} B_{i}\right) \Rightarrow \\
& \frac{\partial \ln L}{\partial a}=\sum_{i=1}^{n} \frac{1}{f\left(\varepsilon_{i}\right)}\left(\frac{b}{(a+b)^{2}} A_{i}-\frac{b}{(a+b)^{2}} B_{i}\right)=\frac{b}{(a+b)^{2}} \sum_{i=1}^{n} \frac{1}{f\left(\varepsilon_{i}\right)}\left(A_{i}-B_{i}\right), \\
& \frac{\partial \ln L}{\partial b}=\sum_{i=1}^{n} \frac{1}{f\left(\varepsilon_{i}\right)}\left(\frac{-a}{(a+b)^{2}} A_{i}+\frac{a}{(a+b)^{2}} B_{i}\right)=\frac{-a}{(a+b)^{2}} \sum_{i=1}^{n} \frac{1}{f\left(\varepsilon_{i}\right)}\left(A_{i}-B_{i}\right) .
\end{aligned}
$$

It is evident that these two expressions are linearly dependent. So setting the score of the likelihood equal to zero will produce a system of equations that are not linearly independent and so it will not have a unique solution. This means that $a, b$ are not separately identifiable. ${ }^{3}$ Moreover, we will obtain the same expression and consequent firstorder condition as above if we use the mixture coefficient $m$ defined earlier. This is the parameter that is identifiable.

At first sight this may appear to be a serious setback, since it does not allow us to estimate the marginal moments of the one-sided error components. But we are still able to estimate a host of other characteristics of the sample that are more important as regards the characterization of the real-world situation and any policy prescriptions. Specifically, from the moment equations in [4.6] and in [4.7], we see that the parameters $a, b$ affect the moments as location-shifters only. This carries over to the JLMS measures for each observation $i$.

Consequently, we will still be able to rank the observations of the sample based on the $w_{i}$ and $u_{i}$ JLMS estimates we will obtain, which is important in order to detect subsets of the sample where the effects of the one-sided error terms are more pronounced and/or asymmetric, by stratifying the sample and observe the rankings present in the various strata.

Also, as we will see in a while, we are still able to estimate the magnitude of the net effect that the one-sided terms combined exert on the outcome. Policy prescriptions for the issues we are considering are in most cases called for when there exists a market imbalance -

[^26]i.e. when the net effect is away from zero, pushing the observed outcome away from the ideal equilibrium point that is free of inefficiencies. In theory, we would want a market free of inefficiencies. In practice, if the evidence says that any existing inefficiencies cancel out, it would perhaps be best to refrain from intervening and let the market function on its own devices.

Moreover, we will be able to test whether the two one-sided error components are actually dependent or not, and to estimate whether they tend to move in the same direction or not (positive/negative correlation). Both of these aspects are important properties of the real-world situation and can critically inform any policy prescriptions. Finally, combining these two pieces of information, we will be able to obtain in some cases lower and in other cases upper bounds for the parameters $a, b$, and consequently, bounds for $E(w), E(u)$ as well as for the observation-specific JLMS measures.

Using $m$ and defining the shortcuts

$$
\omega_{2} \equiv \frac{\varepsilon}{\sigma_{v}}+b^{\prime} \sigma_{v}, \quad \omega_{3} \equiv \frac{\varepsilon}{\sigma_{v}}-a^{\prime} \sigma_{v},
$$

we can conveniently re-write the density as

$$
\begin{equation*}
f_{\varepsilon}(\varepsilon)=\sqrt{2 \pi} \phi\left(\varepsilon / \sigma_{v}\right)\left[m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)\right] . \tag{4.9}
\end{equation*}
$$

The four parameters $\left(\sigma_{v}, a^{\prime}, b^{\prime}, m\right)$ are now identifiable.

The MGF of the distribution is

$$
\begin{equation*}
M_{\varepsilon}(s)=E(\exp \{s \varepsilon\})=M_{z}(s) M_{v}(s)=\left(\frac{m b^{\prime}}{\left(b^{\prime}+s\right)}+\frac{(1-m) a^{\prime}}{\left(a^{\prime}-s\right)}\right) \exp \left\{\frac{1}{2} \sigma_{v}^{2} s^{2}\right\} \tag{4.10}
\end{equation*}
$$

Its cumulative distribution function is

$$
\begin{align*}
F_{\varepsilon}(\varepsilon)=\Phi\left(\varepsilon / \sigma_{v}\right) & +\frac{a}{a+b} \exp \left\{\frac{1}{2} \sigma_{v}^{2} b^{\prime 2}\right\} \exp \left\{b^{\prime} \varepsilon\right\} \Phi\left(-\frac{\varepsilon}{\sigma_{v}}-\sigma_{v} b^{\prime}\right)  \tag{4.11}\\
& -\frac{b}{a+b} \exp \left\{\frac{1}{2} \sigma_{v}^{2} a^{\prime 2}\right\} \exp \left\{-a^{\prime} \varepsilon\right\} \Phi\left(\frac{\varepsilon}{\sigma_{v}}-\sigma_{v} a^{\prime}\right)
\end{align*}
$$

This can be further re-written as

$$
\begin{equation*}
F_{\varepsilon}(\varepsilon)=\Phi\left(\varepsilon / \sigma_{v}\right)+\sqrt{2 \pi} \cdot \phi\left(\varepsilon / \sigma_{v}\right)\left[m \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)-(1-m) \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)\right] . \tag{4.12}
\end{equation*}
$$

## I.2.2. Skewness of the 2TSF Correlated Exponential distribution.

The sign of skewness of the composite error $\varepsilon, \operatorname{sign}\left\{\gamma_{1}(\varepsilon)\right\}$, depends on the sign of its third cumulant $\kappa_{3}(\varepsilon)$ which in turns depends on the sign of the third cumulant of $z=w-u, \kappa_{3}(z)$, since, under the assumption that $v$ follows a zero-mean normal independent of $z$ we have $\kappa_{3}(\varepsilon)=\kappa_{3}(v)+\kappa_{3}(z)=\kappa_{3}(z)$, since $\kappa_{3}(v)=0$.

With independent one-sided error terms, we have that $\kappa_{3}(z)=\kappa_{3}(w)-\kappa_{3}(u)$, and we have shown in chapter 3 that the third cumulants of $w$ and of $u$, when they are either Halfnormal or Exponential random variables, are a monotonic increasing function of their mean, so in these cases we will have $\kappa_{3}(w)-\kappa_{3}(u)<(>) 0 \Leftrightarrow E(w)-E(u)<(>) 0$ and so that $\operatorname{sign}\left\{\gamma_{1}(\varepsilon)\right\}=\operatorname{sign}\{E(w)-E(u)\}$.

But if the one-sided terms are in reality correlated, then it may be the case that we have $\kappa_{3}(\varepsilon)<0$ while at the same time $E(w)-E(u)>0$. So, if we ignore the possibility of dependence while it exists, we may obtain misleading results.

For the 2TSF Correlated Exponential error, we have that

$$
\begin{align*}
\kappa_{3}(\varepsilon)=\kappa_{3}(w-u)=\kappa_{3}(z) & =E\left(z^{3}\right)-3 E(z) E\left(z^{2}\right)+2[E(z)]^{3} \\
& =E\left(z^{3}\right)-E(z)\left[E\left(z^{2}\right)+2 \operatorname{Var}(z)\right] \tag{4.13}
\end{align*}
$$

The term in brackets is positive. So if $E\left(z^{3}\right)<0$ and $E(z)>0$ we will certainly obtain $\kappa_{3}(\varepsilon)<0$ and negative skewness, while if $E\left(z^{3}\right)>0$ and $E(z)<0$ we will certainly obtain $\kappa_{3}(\varepsilon)>0$ and positive skewness. In both cases, the sign of skewness of $\varepsilon$ will be the opposite of the sign of $E(z)=E(w)-E(u)$. These two cases will happen whenever any one of the following relations between the parameters hold:

Table 1: Values of parameters and sign of moments.

| Parameters relation | Sign of moments and cumulants |
| :---: | :---: |
| $\frac{m}{1-m}<\frac{b^{\prime}}{a^{\prime}}<\left(\frac{m}{1-m}\right)^{1 / 3}$ | $E(z)>0, E\left(z^{3}\right)<0, \kappa_{3}(\varepsilon)<0$ |
| $\left(\frac{m}{1-m}\right)^{1 / 3}<\frac{b^{\prime}}{a^{\prime}}<\frac{m}{1-m}$ | $E(z)<0, E\left(z^{3}\right)>0, \kappa_{3}(\varepsilon)>0$ |

The above affect also the symmetric case: a non-symmetric composite error may coexist with $E(w)=E(u)$, or we may have zero skewness together with $E(w) \neq E(u)$ (in which case $\varepsilon$ will be symmetric but not around zero). These results can be seen as the 2TSF analogue of those obtained in Smith (2008) for a single-tier SF production model, where if the symmetric normal error component and the negative one-sided term are dependent, we may have positive skewness of the composite error while its mean is negative. In the single-tier framework, this relates to the "wrong skewness" issue (which we will examine more closely in chapter 6). In the 2TSF framework, it highlights the perils of ignoring the possible existence of dependence. For if such dependence exists and we wrongly apply a 2TSF specification with independence, the means of the one-sided terms will necessarily be estimated so as the sign of their difference, the all-important net effect, conforms with the
sign of the skewness, while the true sign of this difference may be the opposite -and this is a misleading estimation result as serious as one can get.

## I.3. Testing for the existence of dependence.

Freund's Bivariate Exponential extension nests independence when $a=a^{\prime}$ and $b=b^{\prime}$. This implies that if the two one-sided components are independent we will have

$$
\begin{equation*}
m \equiv \frac{a}{a+b}=\frac{a^{\prime}}{a^{\prime}+b^{\prime}} \Rightarrow(1-m) a^{\prime}=m b^{\prime} \tag{4.14}
\end{equation*}
$$

This equality relation is a necessary condition for independence, and it is formed by identifiable parameters. So we can formulate an asymptotically valid statistical test with null hypothesis $\mathrm{H}_{0}:(1-m) a^{\prime}=m b^{\prime}$, which in practice will be based on the difference $T=(1-\hat{m}) \hat{a}^{\prime}-\hat{m} \hat{b}^{\prime}$. If the null hypothesis is rejected we have statistical support for dependence. But if the null hypothesis is not rejected, there is still the possibility that dependence exists, and the test is inconclusive. The test shares the same philosophy and some common characteristics as the well known Hausman (1978) tests (or "vector of contrasts" tests), since it too essentially tests whether two estimators have the same probability limit.

Asymptotically we obtain

$$
\begin{equation*}
\left.\sqrt{n} T\right|_{\mathrm{H}_{0}} \longrightarrow(1-m)\left[\sqrt{n}\left(\hat{a}^{\prime}-a^{\prime}\right)\right]-m\left[\sqrt{n}\left(\hat{b}^{\prime}-b^{\prime}\right)\right]-\left(a^{\prime}+b^{\prime}\right)[\sqrt{n}(\hat{m}-m)] . \tag{4.15}
\end{equation*}
$$

If the null hypothesis $\mathrm{H}_{0}:(1-m) a^{\prime}=m b^{\prime}$ is not correct, the term $\sqrt{n}\left[(1-m) a^{\prime}-m b^{\prime}\right]$ must be added to [4.15], so under the alternative, the value of $\sqrt{n} T$ goes to infinity. Therefore the test is consistent. On the other hand, due to this fact, power studies can only be of a local nature, as in Hausman (1978).

Given the asymptotic properties of the maximum likelihood estimator, at the limit the right-hand side of [4.15] is a scaled sum of three dependent zero-mean Normal random
variables that follow jointly a multivariate Normal distribution. So their sum is a Normal random variable itself. Regarding the limiting variance of the statistic, it does not have a simple form as in the Hausman test case. The simplification there rested on an assumption that one of the two contrasted estimators is efficient while the other was not, under the null. No such property exists in our case, so we cannot avoid the "messy calculations" mentioned by Hausman to obtain the variance (in fact, not that messy, at least in our case). Consistent estimators of all magnitudes involved will be provided by the variance-covariance matrix of the MLE. We obtain

$$
\begin{align*}
\operatorname{Var}(T)=(1-m)^{2} \operatorname{Var} & \left(\hat{a}^{\prime}\right)+m^{2} \operatorname{Var}\left(\hat{b}^{\prime}\right)+\left(a^{\prime}+b^{\prime}\right)^{2} \operatorname{Var}(\hat{m}) \\
& -2(1-m) m \operatorname{Cov}\left(\hat{a}^{\prime}, \hat{b}^{\prime}\right)  \tag{4.16}\\
& -2(1-m)\left(a^{\prime}+b^{\prime}\right) \operatorname{Cov}\left(\hat{a}^{\prime}, \hat{m}\right) \\
& +2\left(a^{\prime}+b^{\prime}\right) m \operatorname{Cov}\left(\hat{m}, \hat{b}^{\prime}\right) .
\end{align*}
$$

Then we obtain a chi-square test,

$$
\begin{equation*}
\hat{q} \equiv \frac{(\sqrt{n} T)^{2}}{\hat{\operatorname{Var}}(\sqrt{n} T)}=\frac{T^{2}}{\hat{\operatorname{Var}}(T)}=\frac{\left((1-\hat{m}) \hat{a}^{\prime}-\hat{m} \hat{b}^{\prime}\right)^{2}}{\hat{\operatorname{Var}}(T)} \xrightarrow[\mathrm{H}_{0}]{d} \chi_{1}^{2} . \tag{4.17}
\end{equation*}
$$

In the Technical Appendix we provide the theoretical derivations related to this test, as well as a simple Monte Carlo simulation that verifies the chi-square distributional result.

## I.4. The distribution of the difference $z=w-u$.

Due to the non-identifiability of the parameters $a, b$, the moments of the marginal distributions are not identifiable, as is clear from [4.6]. In light of this, the distribution of their difference becomes an important source of information. For $z=w-u$ we have the density

$$
f_{Z}(z)= \begin{cases}m b^{\prime} \exp \left\{b^{\prime} z\right\} & z<0  \tag{4.18}\\ (1-m) a^{\prime} \exp \left\{-a^{\prime} z\right\} & z \geq 0\end{cases}
$$

This density has a discontinuity at zero, but this doesn't create any problems since the distribution function integral is convergent, the probability mass is finite for the negative branch also, and it obviously integrates to unity.

Moreover, we see that the distribution of the difference is fully characterized by the three identifiable parameters. Due to the existence of dependence, this density is needed in order to obtain correct net effects, even at the sample level, since, for example, $E\left(e^{z}\right)=E\left(e^{w-u}\right) \neq E\left(e^{w}\right) E\left(e^{-u}\right)$. In fact, its MGF is

$$
\begin{equation*}
M_{z}(s)=E(\exp \{s z\})=\frac{(1-m) a^{\prime}}{\left(a^{\prime}-s\right)}+\frac{m b^{\prime}}{\left(b^{\prime}+s\right)} \tag{4.19}
\end{equation*}
$$

We can also easily calculate its distribution function,

$$
F_{z}(z)= \begin{cases}m \exp \left\{b^{\prime} z\right\} & z<0  \tag{4.20}\\ 1-(1-m) \exp \left\{-a^{\prime} z\right\} & z \geq 0\end{cases}
$$

This is useful for obtaining probabilistic conclusions at sample level. For example we see that

$$
\begin{equation*}
\operatorname{Pr}(w-u>0)=\operatorname{Pr}(z>0)=1-\operatorname{Pr}(z \leq 0)=1-m . \tag{4.21}
\end{equation*}
$$

Also, it is trivial to obtain the quantile function of this distribution, and so calculate quantiles and arrive at a more complete picture.

## I.5. Sample-level expected values and individual JLMS measures.

Due to the identifiability issue, sample-level and individual JLMS measures are not fully computable. But, they are computable up to a location-shift or up to scale, which means that we can obtain the ranking of individual observations based on the JLMS measures.

## I.5.1. Specification in levels.

When the dependent variable enters the specification in levels, we have already provided the unconditional means in eq. [4.6], which are to be used as sample-level metrics. These are not computable, since they include the sum of the non-identifiable parameters, but we can compute their difference

$$
\begin{equation*}
E(z)=E(w-u)=\frac{1-m}{a^{\prime}}-\frac{m}{b^{\prime}}=\frac{(1-m) b^{\prime}-m a^{\prime}}{a^{\prime} b^{\prime}} . \tag{4.22}
\end{equation*}
$$

Later on we will show how we can obtain upper or lower bounds for $E(w), E(u)$.
The variance of $z$, useful in order to assess how much of the variability of the composite error term comes from the non-negative error components, is

$$
\begin{equation*}
\operatorname{Var}(z)=\operatorname{Var}(w)+\operatorname{Var}(u)-2 \operatorname{Cov}(w, u)=\frac{1-m^{2}}{\left(a^{\prime}\right)^{2}}+\frac{m(2-m)}{\left(b^{\prime}\right)^{2}}+\frac{2 m(1-m)}{a^{\prime} b^{\prime}} . \tag{4.23}
\end{equation*}
$$

We can also calculate the difference of the variances

$$
\begin{equation*}
\operatorname{Var}(w)-\operatorname{Var}(u)=\frac{1-m^{2}}{a^{\prime 2}}-\frac{m(2-m)}{b^{\prime 2}}, \tag{4.24}
\end{equation*}
$$

and determine which one is larger.
The corresponding conditional expected values at observation-level per the JLMS approach are

$$
\begin{align*}
& E(w \mid \varepsilon)-\frac{1}{a+b}=\frac{\sigma_{v}(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\}\left[\omega_{3} \Phi\left(\omega_{3}\right)+\phi\left(\omega_{3}\right)\right]}{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)},  \tag{4.25}\\
& E(u \mid \varepsilon)-\frac{1}{a+b}=\frac{\sigma_{v} m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\}\left[\phi\left(\omega_{2}\right)-\omega_{2} \Phi\left(-\omega_{2}\right)\right]}{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)} . \tag{4.26}
\end{align*}
$$

The right-hand sides are computable, and provide us with all the information about the distribution of the conditional expected values, except its exact location. Such a shift does not affect the rankings of individual observations, or the relative allocation of probability mass, or properties like the variance and the skewness of these distributions.

Moreover, evidently we can calculate the difference $E(w \mid \varepsilon)-E(u \mid \varepsilon)$, and obtain the estimated net effect on the outcome for each observation.

## I.5.2. Specification in logarithms.

We turn now to the case when the dependent variable enters the specification in logarithms. Here we need the expected values of the exponentiated variables, which are not computable.

Due to dependence, the net effect is $E\left(e^{w-u}\right) \neq E\left(e^{w}\right) E\left(e^{-u}\right)$. But $E\left(e^{w-u}\right)=E\left(e^{z}\right)$ and by the latter's MGF (eq. [4.19]) we obtain

$$
\begin{equation*}
E\left(e^{w-u}\right)=\frac{m b^{\prime}}{\left(b^{\prime}+1\right)}+\frac{(1-m) a^{\prime}}{\left(a^{\prime}-1\right)} . \tag{4.27}
\end{equation*}
$$

This can also be seen as a convex combination of the MGF's of two Exponential random variables, the one with rate parameter $a^{\prime}$ and MGF evaluated at +1 and the other with rate parameter $b^{\prime}$ and MGF evaluated at -1 .

The corresponding conditional expected values to be used for measures at observation level are

$$
\begin{align*}
& \left(1 \mp \frac{1}{a+b}\right) E(\exp \{ \pm w\} \mid \varepsilon)= \\
& \quad=\frac{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2}\left(\omega_{3} \pm \sigma_{v}\right)^{2}\right\} \Phi\left(\omega_{3} \pm \sigma_{v}\right)}{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)} \tag{4.28}
\end{align*}
$$

$$
\begin{align*}
& \left(1 \mp \frac{1}{a+b}\right) E(\exp \{ \pm u\} \mid \varepsilon)= \\
& \quad=\frac{m b^{\prime} \exp \left\{\frac{1}{2}\left(\omega_{2} \mp \sigma_{v}\right)^{2}\right\} \Phi\left( \pm \sigma_{v}-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \cdot \Phi\left(\omega_{3}\right)}{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)} . \tag{4.29}
\end{align*}
$$

The existence of $E(\exp \{w\} \mid \varepsilon), E(\exp \{u\} \mid \varepsilon)$ requires $a+b>1$, which is not unreasonable, since both parameters essentially function as rate parameters in an Exponential distribution. With both smaller than zero, the distributions would become too fat-tailed. Here too, the right-hand sides are computable, and so we can rank the observations.

Turning to the net effect on the outcome, it is

$$
\begin{align*}
& E(\exp \{w-u\} \mid \varepsilon)= \\
& \quad=\frac{m b^{\prime} \exp \left\{\frac{1}{2}\left(\omega_{2}+\sigma_{v}\right)^{2}\right\} \Phi\left(-\omega_{2}-\sigma_{v}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2}\left(\omega_{3}+\sigma_{v}\right)^{2}\right\} \Phi\left(\omega_{3}+\sigma_{v}\right)}{m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)} . \tag{4.30}
\end{align*}
$$

## I.5.3. Sign of covariance and bounds for the non-identifiable parameters and

 marginal moments.In this section we obtain the sign of the covariance $\operatorname{Cov}(w, u)$. This is of importance on its own, but also, together with the sign of the $T$-statistic from the independence test, permits us to obtain bounds of the non-identified parameters and the marginal moments.

## I.5.3.1 Sign of $\operatorname{Cov}(w, u)$.

The form of eq. [4.25] and [4.26] allows us to obtain the sign of $\operatorname{Cov}(w, u)$ which is an important validation check for the underlying theoretical model, since theoretical considerations prescribe the direction of co-movement of the two one-sided error components.

Using eq. [4.25] and [4.26] we can estimate $\operatorname{Cov}(E(w \mid \varepsilon), E(u \mid \varepsilon))$, since we are missing only a constant factor. Also, we have that $\operatorname{Cov}(w, E(w \mid \varepsilon))>0$ always, and so that

$$
\operatorname{sign}\{\operatorname{Cov}(w, u)\}=\operatorname{sign}\{\operatorname{Cov}(E(w \mid \varepsilon), u)\} .
$$

But
$\operatorname{Cov}(E(w \mid \varepsilon), u)=\operatorname{Cov}\left(E(w \mid \varepsilon), E(u \mid \varepsilon)+e_{u \mid \varepsilon}\right)=\operatorname{Cov}(E(w \mid \varepsilon), E(u \mid \varepsilon))+E\left[E(w \mid \varepsilon) e_{u \mid \varepsilon}\right]$,
where $e_{u \mid \varepsilon}$ is the conditional expectation function error and $E\left(e_{u \mid \varepsilon} \mid \varepsilon\right)=0$. Applying the law of iterated expectations we have
$E\left[E(w \mid \varepsilon) e_{u \mid \varepsilon}\right]=E\left[E\left\{E(w \mid \varepsilon) e_{u \mid \varepsilon} \mid \varepsilon\right\}\right]=E\left[E(w \mid \varepsilon) E\left\{e_{u \mid \varepsilon} \mid \varepsilon\right\}\right]=E[E(w \mid \varepsilon) \cdot 0]=0$.

So we get $\operatorname{Cov}(E(w \mid \varepsilon), u)=\operatorname{Cov}(E(w \mid \varepsilon), E(u \mid \varepsilon))$. It follows that
$\operatorname{sign}\{\operatorname{Cov}(w, u)\}=\operatorname{sign}\{\operatorname{Cov}(E(w \mid \varepsilon), E(u \mid \varepsilon))\}$.

So by computing the right-hand side we can obtain the sign of the left-hand side. ${ }^{4}$

## I.5.3.2. Bounds for the non-identified parameters and the marginal moments.

Assume that for a specific sample we obtain $\hat{\operatorname{Cov}}(w, u)>0$. From the expression for the covariance [4.7] this implies

[^27]$$
\frac{1}{(a+b)^{2}}>\frac{m(1-m)}{a^{\prime} b^{\prime}} \text { but also that } a^{\prime} b^{\prime}>a b
$$
which provides a lower bound for the sum $a+b$ and an upper bound for the product $a b$. Assume also that in the specific sample we obtain
$$
T>0 \Rightarrow(1-m) a^{\prime}>m b^{\prime} \Rightarrow b a^{\prime}>a b^{\prime} \Rightarrow \frac{a^{\prime}}{a}>\frac{b^{\prime}}{b} .
$$

Combined with the previous inequalities, this leads to the relations $a<a^{\prime}, \quad b<\frac{1-m}{m} a^{\prime}$, which are upper bounds for the non-identified parameters. In turn, these provide bounds on the marginal moments of the one-sided error components.

In general we have four different possible combinations of signs of $\operatorname{Cov}(w, u)$ and $T$, and in each case we can obtain information as the above. The Technical Appendix contains the related algebra, here we tabulate the results related to the marginal moments:

Table 2: Sign of covariance and bounds for non-identified parameters.

|  | $\operatorname{Cov}(w, u)>0 \Rightarrow a^{\prime} b^{\prime}>a b$ | $\operatorname{Cov}(w, u)<0 \Rightarrow a^{\prime} b^{\prime}<a b$ |
| :---: | :---: | :---: |
| $T>0 \Rightarrow b a^{\prime}>a b^{\prime}$ | $\begin{gathered} a<a^{\prime}, \quad b<\frac{1-m}{m} a^{\prime} \\ a+b<\min \left\{\frac{a^{\prime}}{m},\left(\frac{a^{\prime} b^{\prime}}{m(1-m)}\right)^{1 / 2}\right\} \end{gathered}$ | $\begin{gathered} a>\frac{m}{1-m} b^{\prime}, \quad b>b^{\prime} \\ a+b>\max \left\{\frac{b^{\prime}}{1-m},\left(\frac{a^{\prime} b^{\prime}}{m(1-m)}\right)^{1 / 2}\right\} \end{gathered}$ |
| $T<0 \Rightarrow b a^{\prime}<a b^{\prime}$ | $\begin{gathered} a<\frac{m}{1-m} b^{\prime}, \quad b<b^{\prime} \\ a+b<\min \left\{\frac{b^{\prime}}{1-m},\left(\frac{a^{\prime} b^{\prime}}{m(1-m)}\right)^{1 / 2}\right\} \end{gathered}$ | $\begin{gathered} a>a^{\prime}, b>\frac{1-m}{m} a^{\prime} \\ a+b>\max \left\{\frac{a^{\prime}}{m},\left(\frac{a^{\prime} b^{\prime}}{m(1-m)}\right)^{1 / 2}\right\} \end{gathered}$ |

The sum of the non-identified parameters $(a+b)$ appears in the various mean and/or variance expressions of both the $w$ and $u$ variables, specifically eq. [4.6], [4.25], [4.26], [4.28],
[4.29], and, depending on which case of the four above we find ourselves, we can obtain upper or lower bounds for them. In certain cases, one can also exploit the sign of the net effect in levels (eq. [4.22])

$$
\operatorname{sign}\{E(w-u)\}=\operatorname{sign}\left\{(1-m) b^{\prime}-m a^{\prime}\right\}=\operatorname{sign}\left\{b b^{\prime}-a a^{\prime}\right\},
$$

to extract even more information.

With this we conclude the presentation of the 2TSF Correlated Exponential specification. We have seen that despite the identification issue, the specification provides all information and metrics of prime interest related to the one-sided error components: their net effect on the dependent variable, which one exhibits greater variability, the ranking of individual observations separately per each one-sided component but also based on the net effect, as well as the direction of the co-movement (sign of covariance) between the two. We will apply this specification in chapter 5 .

We turn now our attention to the issue of regressor endogeneity.

## II. Accounting for regressor endogeneity using Copulas.

In a regression setup, regressor endogeneity due to correlation with the error term can arise for different reasons: measurement error, omitted variables that are correlated with the observables, or a direct distributional assumption on the error term that makes it unlikely to be the conditional expectation function error which is by design mean-independent from the regressors. In the 2TSF framework, we explicitly model unobservable variables, and we make specific distributional assumptions on the error term. Whether the specific unobservables are correlated with the regressors is in principle a case-by-case matter, but it can certainly happen.

Some examples for regressor endogeneity in the 2TSF framework are the following: the labor market has been from the beginning an important market for empirical 2TSF applications, especially the earnings equation. One established determinant of the wage is
years of professional experience. But in many cases, this employee characteristic is not available, and researchers resort to using "potential experience" instead, meaning "maximum possible professional experience", after taking into account the age, and the years of education of the employee. By design, here we have a mismeasured regressor, and we induce correlation between the mismeasured regressor and the error term that now hosts the measurement error also.

As a second example, the two-sided nature of the 2TSF framework reasonably requires regressors that describe both parties in an economic transaction. But such "matched" data sets are not always available, and in many cases we have available data on only one party (again the case of the earnings equation is a clear example, were usually firm-characteristics are absent from the regressor matrix). In such cases, the issue is whether these "omitted variables" are correlated with the available regressors: for the earnings equation, this would require for example that a "self-selection" tendency operates here, with firms attracting subcategories of workers with common attributes that align adequately with the firm's characteristics, but not other subcategories of workers. Further, the argument goes, the firms themselves re-enforce this tendency by eventually hiring those prospects that appear to match the "company's culture". Whether such degree of homogeneity actually characterizes the personnel in a firm is debatable, but the above is certainly a plausible scenario.

A third possible source of regressor endogeneity is the correlation of the regressors with the components of the error term. As an example, consider the Health services market and the "reservation price" 2TSF framework of Gaynor and Polachek (1994, see chapter 2). Here the regressors that represent the buyer (patient) side and determine the "maximum acceptable fee" may include for example data on the severity of the illness, or the income/wealth of the patient. At the same time the error component $w$ represents the priceincreasing effect above full-information equilibrium, due to the "patient's ignorance". We expect that the severity of illness will correlate positively with "patient's ignorance" because it limits the time the patient will consume in searching for alternative providers. We also can argue that a higher income/wealth will correlate positively with "patient's ignorance", because it increases the marginal opportunity cost of time spent searching, and at the same time it decreases the marginal utility of wealth: both aspects create a tendency of less search
and less bargaining with the provider, hence a higher value for $w$ when income/wealth take higher values.

We see that in the 2TSF framework, correlation of regressors with the error term should be expected for various reasons in different circumstances, and so, having a way to handle endogeneity is important. The dominant method in econometrics is the use of Instrumental Variables. The advantage of this approach is that it is free of distributional assumptions and so allows standard least-squares estimation procedures to go through. Its weakness is the ever doubtful assumption that the instruments are valid, namely that they satisfy the heavy requirement that while they are correlated with the endogenous regressor, they are not themselves correlated with the error term.

An estimation approach that can be said to be the "mirror image" of IV-estimation was recently proposed for the 2TSF framework by Parmeter (2017). The author exploited the "Scaling property" that all random variables with a single-parameter distribution possess, and proposed a non-linear least squares (NLS) estimator for the 2TSF model, where the onesided error components are assumed to be functions of other variables for which we have data, and so they are modeled as non-linear functions of these variables alongside the regressors.

If such data are indeed available, the above approach achieves two things at once: first, it dispenses with the need to make distributional assumptions on the error term. And this is because the error term is no longer composite, since the one-sided components are now directly estimable and present in the specification through data series. We do not even need to assume that the remaining error is normally distributed since we are applying NLS. Second, any correlation between the regressors and the one-sided error terms has now become correlation between the elements of an augmented set of regressors, which is fine and in fact sought after, up to a degree. Hence, "regressor endogeneity" is handled not by using instruments instead of the endogenous regressors, but by using observable variables instead of unobservable error components (hence the characterization of the procedure as a "mirror image" of IV estimation). The exploitation of the Scaling Property has similarities with the use of proxy variables, but it is not an identical method, since here we do not proxy the unobservables but rather, we analyze them into their determining factors.

Still, it remains true that the unobservables with which we usually deal in 2TSF models, like incomplete information, bargaining power, or the more exotic ones like selfassessment of life quality or job satisfaction, are such that it won't be easy to find observable explanatory variables for them (although such explanatory variables have certainly already been used in some of the empirical applications of the 2TSF model). Therefore in this section we develop a Copula model using the Gaussian Copula to account for regressor endogeneity, that dispenses with the need to have instruments.

## II.1. The Copula approach.

Copulas have been known for decades in the statistical community, but their use in Econometrics is rather sparse (except in the field of Finance). While the almost necessary reference for an introduction to Copulas is the book by Nelsen (2006), we find much more accessible for a first read to be the paper of Trivedi \& Zimmer (2005) that includes the minimum basic theory, targets specifically econometric applications, and is also a guide to what an econometrician would want to search further. But the authors do not treat explicitly and in detail the issues arising in a regression setup, something that we cover here.

Regarding basic Copula theory, for our purposes here it suffices to ascertain the following facts:

1) A "Copula", denoted by $C$, is a multivariate joint distribution function whose marginals are all uniform $U(0,1)$, so for $m$ random variables it is a $[0,1]^{m} \times[0,1]^{m}$ function.
2) Sklar's theorem (essense) ${ }^{5}$ : Let $X_{i}, i=1, \ldots, m$ be random variables with marginal distribution functions $F_{i}\left(x_{i}\right), i=1, \ldots, m$, and let $H\left(X_{1}, \ldots, X_{m}\right)$ with domain $D^{m}$ be their joint distribution function. Then there exists a Copula such that

$$
H\left(X_{1}, \ldots, X_{m}\right)=C_{X_{1} \cdots X_{m}}\left(F_{1}\left(X_{1}\right), \ldots, F_{m}\left(X_{m}\right)\right),
$$

or

[^28]$$
H\left(x_{1}, \ldots, x_{m}\right)=C_{X_{1} \cdots X_{m}}\left(F_{1}\left(x_{1}\right), \ldots, F_{m}\left(x_{m}\right)\right), \quad \forall\left(x_{1}, \ldots, x_{m}\right) \in D^{m} .
$$

In words, this Copula is an alternative representation of the actual joint distribution of this collection of random variables. When the random variables are continuous, the above Copula is unique. But if the random variables are discrete, there are many Copulas that can represent the joint distribution function, and a host of other issues arise that are detailed in Genest \& Nešlehová (2007). We will assume continuous random variables in what follows, except where noted otherwise. But in many cases, we will have to deal with endogenous regressors that are discrete-valued, or measured on a discrete scale. For such cases, we will apply a simple method to make a discrete random variable continuous while maintaining its probabilistic structure.
3) If the joint density exists, $h\left(x_{1}, \ldots, x_{m}\right)$, and the marginal densities are denoted $f_{i}\left(x_{i}\right), i=1, \ldots, m$, we have

$$
\begin{equation*}
h\left(x_{1}, \ldots, x_{m}\right)=c_{X_{1} \cdots X_{m}}\left(F_{1}\left(x_{1}\right), \ldots, F_{m}\left(x_{m}\right)\right) \cdot \prod_{i=1}^{m} f_{i}\left(x_{i}\right), \tag{4.32}
\end{equation*}
$$

where

$$
c_{X_{1} \cdots X_{m}}\left(F_{1}\left(X_{1}\right), \ldots, F_{m}\left(X_{m}\right)\right)=\frac{\partial^{m} C_{X_{1} \cdots X_{m}}\left(F_{1}\left(X_{1}\right), \ldots, F_{m}\left(X_{m}\right)\right)}{\partial F_{1}\left(X_{1}\right) \cdots \partial F_{m}\left(X_{m}\right)},
$$

is the copula density. Note that we differentiate the Copula with respect to the $F_{i}\left(X_{i}\right)$ terms, which are its arguments.

For maximum likelihood estimation, eq. [4.32] is the crucial result: it shows that through the Copula approach we can separate the dependence structure from the independence joint density (the product of the marginals) in a very convenient multiplicative way which becomes additive under the standard logarithmic transformation.

Note that in the copula density only dependent variables appear. This is immediate from the fact that the Copula, being a distribution function, separates as a distribution
function for any independent variable in the collection. For example, for three variables where the first one is independent from the other two, we have

$$
F\left(X_{1}, X_{2}, X_{3}\right)=F_{1}\left(X_{1}\right) F_{23}\left(X_{2}, X_{3}\right),
$$

and analogously for the Copula

$$
C\left[F_{1}\left(X_{1}\right), F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right]=C_{1}\left[F_{1}\left(X_{1}\right)\right] C_{23}\left[F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right],
$$

and so

$$
\begin{aligned}
& c\left[F_{1}\left(X_{1}\right), F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right]=\frac{\partial^{3} C\left[F_{1}\left(X_{1}\right), F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right]}{\partial F_{1}\left(X_{1}\right) \partial F_{2}\left(X_{2}\right) \partial F_{3}\left(X_{3}\right)} \\
& \\
& =\frac{\partial C_{1}\left[F_{1}\left(X_{1}\right)\right]}{\partial F_{1}\left(X_{1}\right)} \frac{\partial^{2} C_{23}\left[F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right]}{\partial F_{2}\left(X_{2}\right) \partial F_{3}\left(X_{3}\right)} \\
& \Rightarrow c\left[F_{1}\left(X_{1}\right), F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right]=c_{1}\left[F_{1}\left(X_{1}\right)\right] \cdot c_{23}\left[F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right] f\left(x_{1}\right) f\left(x_{2}\right) f\left(x_{3}\right) .
\end{aligned}
$$

But $c_{1}\left[F_{1}\left(X_{1}\right)\right]=1$ since it is the density of a Uniform $U(0,1)$ random variable and so

$$
c\left[F_{1}\left(X_{1}\right), F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right]=c_{23}\left[F_{2}\left(X_{2}\right), F_{3}\left(X_{3}\right)\right] f\left(x_{1}\right) f\left(x_{2}\right) f\left(x_{3}\right) .
$$

In order to account for regressor endogeneity, what we do is to formulate a model based on the joint density only of the endogenous regressors and the error term, and assume a functional form for the copula density. Compared to the Instrumental Variables method, the Copula approach solves a statistical issue (inconsistency due to endogeneity) by using a statistical solution: it directly recognizes and models the statistical dependence, rather than trying to bypass it by invoking a stretched combination of structural relations as in the Instrumental Variables method.

## II.2. Estimation of a regression model with the Gaussian Copula.

We are examining a cross-sectional i.i.d. linear regression model with a 2TSF composite error,

$$
\begin{equation*}
y_{i}=\mathbf{x}_{i}^{\prime} \beta+\varepsilon_{i}, \quad \varepsilon_{i}=v_{i}+w_{i}-u_{i}, i=1, \ldots, n . \tag{4.33}
\end{equation*}
$$

In fact the approach we will describe handles any assumption on the error term, as well as non-linear specifications. The approach takes its lead from Tran \& Tsionas (2015) who described estimation of a single-tier SF model with a Gaussian Copula for the case of a single endogenous regressor, but we deal also with issues that arise when we want to consider more than one endogenous regressor, as well as discrete regressors and regressors that are deterministic functions of other regressors. We discuss the merits of choosing the Gaussian Copula after presenting the model.

Amsler Prokhorov \& Schmidt (2015) considered also the use of the Gaussian copula to account for heterogeneity in an SF model. But they adopted a "control function" approach, where instruments are available for the endogenous regressors, although they do not make them uncorrelated with the error term. Also, their approach by design requires simulations and numerical estimations, while in what we will develop here numerical estimation will be needed only if we don't make an assumption on the distribution of the composite error term.

The model has $K+1$ regressors, including the constant term. For some of the nonconstant regressors, say the first $m$ of them, we have reasons to believe that they are endogenous, i.e. correlated with the error term (of the same observation). We need to model only the joint distribution function of the endogenous regressors and the error term for observation $i$, which is written in copula form

$$
\begin{equation*}
H\left(X_{1 i}, \ldots, X_{m i}, \varepsilon_{i}\right)=C\left[F_{1}\left(X_{1 i}\right), \ldots, F_{m}\left(X_{m i}\right), F_{\varepsilon}\left(\varepsilon_{i}\right)\right] . \tag{4.34}
\end{equation*}
$$

So the joint density here will be

$$
\begin{equation*}
h\left(X_{1 i}, \ldots, X_{m i}, \varepsilon_{i}\right)=c\left[F_{1}\left(X_{1 i}\right), \ldots, F_{m}\left(X_{m i}\right), F_{\varepsilon}\left(\varepsilon_{i}\right)\right] \cdot\left(\prod_{j=1}^{m} f_{j}\left(x_{j i}\right)\right) \cdot f_{\varepsilon}\left(\varepsilon_{i}\right) \tag{4.35}
\end{equation*}
$$

and the log-likelihood for observation $i$ is

$$
\begin{equation*}
\ell_{i}=\ln c\left[F_{1}\left(x_{1 i}\right), \ldots, F_{m}\left(x_{m i}\right), F_{\varepsilon}\left(\varepsilon_{i}\right)\right]+\ln \left(\prod_{j=1}^{m} f_{j}\left(x_{j i}\right)\right)+\ln f_{\varepsilon}\left(\varepsilon_{i}\right) \tag{4.36}
\end{equation*}
$$

Note that the main parameters of interest, namely the beta coefficients, the parameters of the error distribution, as well as the dependence parameters included in the copula density, do not appear in the second component of the likelihood, that has the marginal densities of the regressors, so we can ignore that term in maximum likelihood estimation.

Moreover, the distribution functions of the endogenous regressors do not contain parameters of interest, and can be estimated non-parametrically. The standard way to do this is by the empirical distribution function,

$$
\begin{equation*}
\breve{x}_{j i} \equiv \hat{F}_{j}\left(x_{j i}\right)=\frac{1}{n+1} \sum_{k=1}^{n} I\left\{x_{j k} \leq x_{j i}\right\}, \quad i=1, \ldots, n, \quad j=1, \ldots, m . \tag{4.37}
\end{equation*}
$$

Note that here we need to obtain an estimated probability for each observation, we do not form the empirical distribution function itself in the usual way (which would group identical or nearly identical observations to form representative quantiles). So we obtain a series that has the length of the sample. The term $1 /(n+1)$ instead of $1 / n$ is used in order to avoid difficulties that may arise from unboundedness of the copula density as some empirical probabilities tend to unity.

## II.2.1. Dealing with discrete and/or discretely-measured variables.

An alternative and more robust approach to calculate the probability series for the endogenous regressors is to use the concept of the "mid-distribution function",

$$
\begin{equation*}
F_{\text {mid }}(x)=F(x)-\frac{1}{2} P(X=x) . \tag{4.38}
\end{equation*}
$$

The mid-distribution function has been proposed by Parzen (see for example his 2004 paper), to deal with certain problems related to the limiting behavior of sample quantiles of discrete distributions. Specifically, for discrete random variables, if quantiles are defined in the standard way, their limiting distribution is discrete and non-normal, especially when the sample contains ties. On the contrary, sample quantiles defined and obtained by the use of the mid-distribution function have an asymptotically normal distribution (see Ma, Genton \& Parzen 2011 for more on the subject).

Here, we calculate

$$
\begin{equation*}
\breve{x}_{j i}=\hat{F}_{j}\left(x_{j i}\right)=\frac{1}{n+1} \sum_{k=1}^{n} I\left\{x_{j k}<x_{j i}\right\}+\frac{1}{n+1} \sum_{k=1}^{n} \frac{1}{2} I\left\{x_{j k}=x_{j i}\right\} . \tag{4.39}
\end{equation*}
$$

In practice what we do in [4.39] is to not count half of any ties for each observation.
The reason to use the mid-distribution function also for conceptually continuous random variables, is the fact that the data in a sample are always discrete. For example in labor economics, variables like Age, Experience, Education, Tenure, are usually measured/recorded in whole years.

But the use of the empirical mid-distribution function may not be enough to obtain a probability series that behaves like a Uniform random variable, especially when the regressor involved takes on very few values. This is because if $X_{d}$ is a discrete random variable with distribution function $F_{d}\left(x_{d}\right)=\operatorname{Pr}\left(X_{d} \leq x_{d}\right)$, the random variable $F_{d}\left(X_{d}\right)$ does not follow a Uniform $U(0,1)$ distribution. Worse, whatever $X_{d}$ "is", if its sampled values resemble enough a discrete random variable, then the transformed series will not "resemble enough" a Uniform $U(0,1)$. To overcome this problem, we can apply the "continuation transformation" as proposed in Genest, Nešlehová \& Rémillard (2014), Proposition 2.2.

Specifically, let $U$ be a Uniform $U(0,1)$ random variable, independent from an integer-valued random variable $X_{d}$. Then the random variable $X_{c}=X_{d}+U-1$ is continuous. Consider now any value from the support of $X_{d}$, say $x_{d}$, which is an integer and certainly belongs also to the support of $X_{c}$. We have

$$
\operatorname{Pr}\left(X_{c} \leq x_{d}\right)=\operatorname{Pr}\left(X_{d}+U-1 \leq x_{d}\right)=\operatorname{Pr}\left(X_{d} \leq x_{d}+(1-U)\right) .
$$

Since, almost surely, $0<(1-U)<1$, and $X_{d}$ takes only the values $\left\{\ldots, x_{d}, x_{d}+1, \ldots\right\}$ it follows that $\operatorname{Pr}\left(X_{d} \leq x_{d}+(1-U)\right)=\operatorname{Pr}\left(X_{d} \leq x_{d}\right)$ and so that

$$
\begin{equation*}
\operatorname{Pr}\left(X_{c} \leq x_{d}\right)=\operatorname{Pr}\left(X_{d} \leq x_{d}\right) . \tag{4.40}
\end{equation*}
$$

This implies that the probabilities that the distribution function of $X_{c}$ gives for the values in the support of $X_{d}$, are identical with the probabilities provided by the distribution function of $X_{d}$ itself. So if we estimate the empirical distribution function of $X_{c}$, it will "include" also the empirical distribution function of $X_{d}$, as the estimated probabilities at the integer values:

$$
\begin{equation*}
\hat{F}_{Y}\left(x_{d}\right)=\frac{1}{n+1} \sum_{k=1}^{n} I\left\{x_{c, k} \leq x_{d}\right\}=\frac{1}{n+1} \sum_{k=1}^{n} I\left\{x_{d, k} \leq x_{d}\right\}=\hat{F}_{X_{d}}\left(x_{d}\right), \forall x_{d} \in S_{X_{d}} . \tag{4.41}
\end{equation*}
$$

Of course by estimating the empirical distribution function of $X_{c}$ we will obtain many more probability estimates, since the transformed series $X_{c}$ will include also non-integer values. Especially if we do it as in [4.37] or [4.39], where we do not group identical values, but obtain a probability estimate for each point in the sample.

But as is shown formally in Genest et al. (2014), the Copula involving the $X_{c}$-type variables and their distribution functions, is a Copula also for the linked discrete-valued $X_{d}$-type variables (the authors call it the "multilinear extension Copula"). Then we can
validly use it and, moreover, since it now contains only continuous random variables, it has a copula density. ${ }^{6}$

In practice, and since we will be implementing the Gaussian copula, it is good policy to run univariate normality tests on all the variables $\Phi^{-1}\left(\breve{x}_{j}\right)$ where $\Phi$ is the univariate standard Normal distribution function. For those that fail the test, we generate a series of i.i.d. realizations from a Uniform $U(0,1)$ random variable of length $n$, then form the series $X_{c, i}=X_{d, i}+U_{i}-1, i=1, \ldots, n$ and then calculate

$$
\begin{equation*}
\breve{x}_{c, i}=\frac{1}{n+1} \sum_{k=1}^{n} I\left\{x_{c, k} \leq x_{c, i}\right\}, \quad i=1, \ldots, n, \tag{4.42}
\end{equation*}
$$

(or the mid-distribution function alternative, although we expect that there will be no ties). The variable $\breve{x}_{c}$ will be distributed $U(0,1)$ and we can use $\breve{x}_{c}$ in place of $\breve{x}_{d}$. Note that if we have to apply this transformation to more than one regressor, then we should generate and use different and independent series of $n$ realizations from a $U(0,1)$ distribution for each regressor, otherwise we will artificially create dependence (or alter the actually existing one) between the continuation-transformed regressors.

A final detail must also be stressed: suppose we have panel data, say, over a cross section for a number of years, and we want to estimate separate cross-sectional regressions per year. Then the transformation of the regressors for the Copula and any continuationtransformation must be applied per sub-sample separately. If we first apply it in the whole sample and then subsample per year, there is no guarantee that the transformed sub-series will exhibit the properties of a standard Normal random variable. The same principle applies to the case of a cross-sectional sample that we want to break and perform separate estimations on its subsamples: first separate the strata, and then apply the continuation transformation in each stratum.

[^29]
## II.2.2. The likelihood function with a Gaussian copula density.

Having the empirical probabilities of the regressors, we can concentrate on the loglikelihood

$$
\begin{equation*}
\breve{\ell}_{i}=\ln c\left[\breve{x}_{1 i}, \ldots, \breve{x}_{m i}, F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right) ; \mathrm{R}\right]+\ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right), \tag{4.43}
\end{equation*}
$$

where we have explicitly brought onto the surface the various parameters that need to be estimated. $\boldsymbol{\theta}$ holds the parameters of the distribution of the error term, while R holds the parameters of the copula density. Note that since we treat here not just the bivariate copula case but the multivariate one, namely, the case where more than one regressor is considered correlated with the error term, R will include also parameters that reflect dependence between the (transformed) regressors. These are nuisance parameters of no interest, and they can be consistently calculated from the sample directly, since we have data available. This is beneficial because it reduces the number of unknown coefficients to be estimated by the maximum likelihood estimator. We denote $\breve{\mathrm{R}}$ the R matrix with some elements as fixed constants and we further concentrate the log-likelihood into

$$
\begin{equation*}
\breve{\ell}_{i}=\ln c\left[\breve{x}_{1 i}, \ldots, \breve{x}_{m i}, F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right) ; \breve{\mathrm{R}}\right]+\ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right) . \tag{4.44}
\end{equation*}
$$

We now assume the Gaussian form for the copula density. Set

$$
\begin{equation*}
\mathbf{q}_{i}=\left(\Phi^{-1}\left(\breve{x}_{1 i}\right), \ldots, \Phi^{-1}\left(\breve{x}_{m i}\right), \Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right)^{\prime} . \tag{4.45}
\end{equation*}
$$

Note that the first $m$ elements of $\mathbf{q}_{i}$ are fixed numbers (possibly different for each observation), and only the $(m+1)$ - th position includes unknown parameters. The Gaussian copula density is

$$
\begin{align*}
c_{i}^{G}=\operatorname{det}(\mathrm{R})^{-1 / 2} & \cdot \exp \left\{-\frac{1}{2} \mathbf{q}_{i}^{\prime}\left(\mathrm{R}^{-1}-I_{m+1}\right) \mathbf{q}_{i}\right\}  \tag{4.46}\\
& \Rightarrow \ln c_{i}^{G}=-\frac{1}{2} \ln \operatorname{det}(\mathrm{R})-\frac{1}{2} \mathbf{q}_{i}^{\prime}\left(\mathrm{R}^{-1}-I_{m+1}\right) \mathbf{q}_{i}
\end{align*}
$$

where R is the correlation matrix,

$$
\mathrm{R}=\left[\begin{array}{cccc}
1 & \rho_{12} & \cdots & \rho_{1 \varepsilon}  \tag{4.47}\\
\rho_{12} & 1 & & \vdots \\
\vdots & & 1 & \rho_{m \varepsilon} \\
\rho_{1 \varepsilon} & \cdots & \rho_{m \varepsilon} & 1
\end{array}\right]
$$

and where, for example, $\rho_{12}=\operatorname{corr}\left(\Phi^{-1}\left(\breve{x}_{1}\right), \Phi^{-1}\left(\breve{x}_{2}\right)\right)$ : as already said the correlation between the doubly-transformed regressors will be estimated from the sample so we will work with

$$
\breve{\mathrm{R}}=\left[\begin{array}{ccccc}
1 & \hat{\rho}_{12} & \cdots & \hat{\rho}_{1 m} & \rho_{1 \varepsilon}  \tag{4.48}\\
\hat{\rho}_{12} & 1 & \cdots & \hat{\rho}_{2 m} & \rho_{2 \varepsilon} \\
\vdots & \vdots & \ddots & & \vdots \\
\hat{\rho}_{1 m} & \vdots & & \ddots & \rho_{m s} \\
\rho_{1 \varepsilon} & \cdots & \cdots & \rho_{m \varepsilon} & 1
\end{array}\right]
$$

Now the log-likelihood becomes

$$
\begin{equation*}
\breve{\ell}_{i}=-\frac{1}{2} \ln \operatorname{det}(\breve{\mathrm{R}})-\frac{1}{2} \mathbf{q}_{i}^{\prime}\left(\breve{\mathrm{R}}^{-1}-I_{m+1}\right) \mathbf{q}_{i}+\ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right) . \tag{4.49}
\end{equation*}
$$

This can be further simplified since $\mathbf{q}_{i}^{\prime}\left(\breve{\mathrm{R}}^{-1}-I_{m+1}\right) \mathbf{q}_{i}=\mathbf{q}_{i}^{\prime} \breve{\mathrm{R}}^{-1} \mathbf{q}_{i}-\mathbf{q}_{i}^{\prime} \mathbf{q}_{i}$ and the inner product $\mathbf{q}_{i}^{\prime} \mathbf{q}_{i}$ contains unknown parameters only in its last element, which is $\left[\Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right]^{2}$.

So the effective log-likelihood at observation $i$ can be written
$\breve{\ell}_{i}=-\frac{1}{2} \ln \operatorname{det}(\breve{\mathrm{R}})-\frac{1}{2} \mathbf{q}_{i}^{\prime} \breve{\mathrm{R}}^{-1} \mathbf{q}_{i}+\frac{1}{2}\left[\Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right]^{2}+\ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)$,
and the log-likelihood of the sample to be used in maximum likelihood estimation becomes

$$
\begin{align*}
\ln \breve{L}=-\frac{n}{2} \ln \operatorname{det}(\breve{\mathrm{R}})- & \frac{1}{2} \sum_{i=1}^{n} \mathbf{q}_{i}^{\prime} \breve{\mathrm{R}}^{-1} \mathbf{q}_{i}  \tag{4.51}\\
& +\frac{1}{2} \sum_{i=1}^{n}\left[\Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right]^{2}+\sum_{i=1}^{n} \ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right),
\end{align*}
$$

to be maximized over $\left(\breve{\boldsymbol{p}}_{x \varepsilon}, \boldsymbol{\theta}, \beta\right)$, $\breve{\mathbf{p}}_{x \varepsilon}=\left(\rho_{1 \varepsilon}, \ldots, \rho_{m \varepsilon}\right)$.
For the bivariate case, where there is a single dependence parameter, namely the correlation coefficient between the single endogenous regressor and the error term (their transformed counterparts), we have $\breve{R}=R$,

$$
\operatorname{det}(\mathrm{R})=1-\rho_{1 \varepsilon}^{2}, \quad \mathrm{R}^{-1}=\left[\begin{array}{cc}
\frac{1}{1-\rho_{1 \varepsilon}^{2}} & \frac{-\rho_{1 \varepsilon}}{1-\rho_{1 \varepsilon}^{2}} \\
\frac{-\rho_{1 \varepsilon}}{1-\rho_{1 \varepsilon}^{2}} & \frac{1}{1-\rho_{1 \varepsilon}^{2}}
\end{array}\right]
$$

and the effective likelihood at the observation level can be explicitly written

$$
\begin{align*}
\breve{\ell}_{i}=-\frac{1}{2} \ln \left(1-\rho_{1 \varepsilon}^{2}\right)- & \frac{\left[\Phi^{-1}\left(\breve{x}_{1 i}\right)\right]^{2}}{2\left(1-\rho_{1 \varepsilon}^{2}\right)}+\frac{\rho_{1 \varepsilon}}{\left(1-\rho_{1 \varepsilon}^{2}\right)} \Phi^{-1}\left(\breve{x}_{1 i}\right) \Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right) \\
& -\frac{\rho_{1 \varepsilon}^{2}}{2\left(1-\rho_{1 \varepsilon}^{2}\right)}\left[\Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right]^{2}+\ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right) . \tag{4.52}
\end{align*}
$$

Note that the regressor series appear in the likelihood both in their original form as well as transformed for the needs of the copula density.

## II.2.3. Assessing the fit and model selection.

Discarding from the likelihood terms that do not include unknown parameters may be convenient computationally, but we lose the ability to execute likelihood ratio tests or use Information criteria like AIC and BIC to assess the fit between alternative models, in cases where the alternative models are characterized by different sets of regressors or different assumptions about the error distribution, or the Copula. In these cases, for model selection tests the likelihood in [4.49]. And if the competing models have different sets of regressors, we need to go back to [4.36] and estimate empirically the marginal densities of the regressors also

## II.2.4. Probabilistically redundant regressors: squares and interaction terms.

There are cases were some of the regressors are strictly monotonic deterministic functions of another regressor, say $x_{j}=h\left(x_{k}\right)$. For example, it is usual practice to include the square of a positive variable in a regression specification alongside this variable's level, in order to capture a possible non-linear relationship between the regressor and the dependent variable. Since these are monotonic transformations, it follows that the series obtained for $x_{k}$ and $x_{j}$ through the empirical distribution functions will be exactly the same, since the original transformation does not affect the rank/order of each observation. Then, we will also have $\Phi^{-1}\left(\breve{x}_{j i}\right)=\Phi^{-1}\left(\breve{x}_{k i}\right), \forall i$, because $\Phi^{-1}(\cdot)$ is also strictly monotonic. From a theoretical perspective, including both series in the Copula would cause the domain of the Copula distribution to "lose one dimension", to become a ( $n-1$ )-dimensional object in a $n$ dimensional space, resulting in its Lebesgue measure being zero. For example, with two regressors related as the example above, and the error term, the domain of the Copula is an area in $\mathbb{R}^{3}$, it has no volume. This implies that the copula density is not defined. In practice, attempting to do so in an estimation procedure will result in a "perfect colinearity" message. So in cases where an endogenous regressor appears more than once in the regressor matrix,
transformed in a strictly monotonic fashion, we should only include in the copula density the regressor's basic incarnation.

Another frequent occurrence in empirical studies is the inclusion of products of regressors, either selectively so as to capture targeted interaction effects, or comprehensively as an integral part of flexible functional forms like the translog. These regressor products are also to be treated as probabilistically redundant: if the individual elements of the product are considered endogenous, then it is the original variables that have to be included in the copula density specification.

## II.2.5. Estimation steps.

The specific steps of the whole estimation procedure using the Gaussian copula are summarized below:

1. Determine which $m$ regressors are to be treated as endogenous.
2. Calculate the transformed data series for these regressors, first by computing the empirical (mid-)distribution function for each point in the sample for them, $\breve{x}_{j i}$, and then by passing the values obtained through the inverse standard normal distribution function to obtain $\Phi^{-1}\left(\breve{x}_{j i}\right)$.
3. Perform univariate normality tests on each of the $\Phi^{-1}\left(\breve{x}_{j}\right)$ series. For those who fail, re-calculate $\breve{x}_{j}$ by applying the "continuation transformation" described earlier.
4. Calculate the sample correlation coefficients between the final $\Phi^{-1}\left(\breve{x}_{j}\right)$ data series.
5. Construct the $\mathbf{q}_{i}$ vectors and the $\breve{\mathrm{R}}$ matrix (part numbers, part unknown coefficients), and form the log likelihood function, using the assumed distributional specification for the composite error term, with unknown parameters $\left(\breve{\boldsymbol{\rho}}_{x \varepsilon}, \boldsymbol{\theta}, \beta\right)$.
6. Starting values for the ML algorithm are always an issue. Calculate OLS estimates to obtain starting values for the beta coefficients, and the variance of the composite error term (which may provide some guidance for reasonable starting values for the $\boldsymbol{\theta}$ vector). Regarding the correlations between the transformed variables and the transformed error term, theory and other out-of-sample information should be able to determine at least the
sign of these coefficients. In any case, estimation should be executed with different sets of starting values, in order to assess the robustness of the results.

We note finally that for some specifications the distribution function of the composite error term $F_{\varepsilon}(\varepsilon)$ may not be available, in which case numerical estimation is required, see Tran and Tsionas (2015). For the 2TSF specifications that have been presented in this dissertation, we have closed-form distribution functions for the Exponential and Generalized Exponential specifications from chapter 3, and for the Correlated Exponential from the previous section ("closed-form" in the sense of including integrals that are widely implemented as special functions in software programs).

## II.3. Properties of the maximum likelihood estimator, Copula misspecification and the case for the Gaussian Copula.

The model we have presented in the previous section uses the concentrated likelihood [4.51]. Subject to correct specification, Genest, Ghoudi \& Rivest (1995) showed that the MLE for the dependence parameters $\breve{\boldsymbol{\rho}}_{x \varepsilon}$ is consistent and asymptotically normal with the concentrated likelihood also. But we should acknowledge that in most cases we will be misspecifying the copula density.

We apply copula-modeling because we do not want to ignore the possibility of endogenous regressors, and the detrimental effects on the estimation results if the model is misspecified in that respect. But the true Copula of the variables treated as correlated with the error term is also unknown, and any choice of Copula and the corresponding copula density most likely won't be the correct one. So it seems that we have exchanged one form of misspecification for another. Can we say that, nevertheless, we are in a better position? And is the choice of the Gaussian Copula a wise one?

We believe the answers to these two questions are both in the affirmative, and we lay down some arguments to support this.

By "ignoring dependence", in reality we impose on the estimator a very specific assumption: that all dependence parameters are equal to zero. The estimator by design will have to operate under this constraint. On the other hand by including a copula densitys we
allow the estimator to be able to detect the possible dependence, even though it may be forced to do it through a misspecified function. Then, if endogeneity exists, at least part of its effects will be reflected in the estimates of the dependence parameters, reducing the bias of the estimates of the other parameters in the model. In one sentence, it is better to misspecify endogeneity than to ignore it.

## II.3.1. The choice of the Gaussian Copula.

Regarding the choice of the Gaussian Copula specifically, the following arguments can be advanced:

1) It allows for the full $(-1,1)$ range of Pearson's correlation coefficient, something that is not true for Copulas in general. Moreover, as we show later, the estimated correlation coefficients (of the transformed variables) equal the "maximum correlation coefficient" of the original variables.
2) It nests the "Independence Copula", where in reality the copula density is equal to unity and the regressors are not correlated with the error term. Again, this is not a universal feature of Copulas, since some of them do not nest the independence case.
3) It is a radially symmetric Copula and it is for this class of Copulas that we have the single known to us consistency (i.e. robustness) result for the quasi-maximum likelihood estimator in the presence of Copulas (see Prokhorov \& Schmidt 2009, Theorem 5, p.99). ${ }^{7}$
4) It is an elliptical Copula, and this is a desirable feature related to the number of parameters, simulation functionality, etc (see Danaher \& Smith 2011). Another important advantage of elliptical Copulas is that they extend naturally to more than two variables. This is a very useful property since in many cases we expect that more than one regressor will be suspected for endogeneity. With two regressors and the error term, we need a trivariate Copula. ${ }^{8}$

[^30]5) There exists Monte Carlo evidence that the Gaussian Copula is robust against misspecification, related either to the true Copula or to the true distribution of the error term, (see the results of the simulation study in Park \& Gupta 2012).
6) We can partially test the appropriateness of using the Gaussian copula, and even get a sense of how much the data deviate from this assumption. We elaborate on this matter on the next subsection.

## II.4. Testing the validity of the Gaussian Copula.

Whether the Gaussian Copula can be considered the correct specification or not, rests on whether the endogenous regressors, transformed into standard Normal random variables, follow a multivariate Normal distribution (MVN), jointly with the transformed error term. This comes from the following result:

Lemma 4.1: Let $X_{i}, i=1, \ldots, m$ be continuous random variables with marginal distribution functions $F_{i}\left(x_{i}\right), i=1, \ldots, m$, and Copula $C_{X_{1} \cdots X_{m}}\left(F_{1}\left(X_{1}\right), \ldots, F_{m}\left(X_{m}\right)\right)$. Let $\Phi$ be the standard normal distribution function, and $\Phi^{-1}$ its inverse. Let $\Phi_{m}$ be the multivariate standard normal distribution function of dimension $m$.

Then: if the random variables $\Phi^{-1}\left(F_{i}\left(X_{i}\right)\right), i=1, \ldots, m$ follow jointly a Multivariate Normal (MVN) distribution, it holds that

$$
\begin{equation*}
C_{X_{1} \cdots X_{m}}\left(F_{1}\left(X_{1}\right), \ldots, F_{m}\left(X_{m}\right)\right)=\Phi_{m}\left(\Phi^{-1}\left(F_{1}\left(X_{1}\right)\right), \ldots, \Phi^{-1}\left(F_{m}\left(X_{m}\right)\right)\right) . \tag{4.53}
\end{equation*}
$$

The proof of the Lemma can be found in the Technical Appendix. This result clarifies that when using the Gaussian Copula, the existence of "Copula misspecification" is not related to the transformation of the original regressors in any way, or to their marginal distributions, but rests solely on whether the transformed variables follow jointly a MVN distribution or not. They are already standard Normals, but since we expect them to be correlated, as regressors usually are, they may not follow the MVN distribution.

The advantage of assuming the Gaussian Copula then is that we can (partially) test by a formal statistical test for Copula misspecification by testing for Multivariate Normality of
the transformed variables, since a well-developed arsenal of such tests exists. Henze (2002) contains a critical review and presentation of many statistical tests for MVN, while various software programs include ready-made tests for multivariate normality.

The test is "partial", because in our regression setting, we can test for MVN only the collection of the standard Normals corresponding to the regressors, excluding the one corresponding to the error term, $\Phi^{-1}\left(F_{\varepsilon}(\varepsilon)\right)$, since we do not have data on it. But testing for MVN on the rest lends a degree of support to the Gaussian Copula specification, if the null hypothesis is not rejected. In certain cases, even if MVN is rejected, we can even assess whether the departure from multivariate normality is serious enough or not, by using visual inspection tools and plots, and/or perform an outlier/influential observations analysis.

Of course, even if MVN is supported for the transformed regressors, that doesn't guarantee that the addition of the $\Phi^{-1}\left(F_{\varepsilon}(\varepsilon)\right)$ variable in the collection will preserve multivariate normality. Still, the actual consequences of misspecification have to do with the degree to which our modeling assumptions deviate from the true structure. Therefore, adding the MVN-testing procedure to the whole estimation strategy and assessing the degree of deviation has merit. As a final check, after estimating the model one could re-test for MVN including also the residual series from the estimation (after transforming them of course).

## II.5. Interpreting the correlation measures.

Estimating the regression Copula model, we will obtain estimates for $\hat{\rho}_{j \varepsilon}=\operatorname{corr}\left(\Phi^{-1}\left(F_{j}\left(x_{j}\right)\right), \Phi^{-1}\left(F_{\varepsilon}(\varepsilon)\right)\right)$. These are rank correlation coefficients of the "van der Waerden type" (see Hajek, Sidak and Sen 1999, ch. 4), and, as Klaassen and Wellner (1997) show, under joint normality their absolute values estimate consistently and efficiently the "maximum correlation coefficient" of $\left(X_{j}, \varepsilon\right)$,

$$
\begin{equation*}
\left|\hat{\rho}_{j \varepsilon}\right| \xrightarrow{p} \rho_{M}\left(X_{j}, \varepsilon\right) \equiv \sup _{h, g}\left\{\rho\left[h\left(X_{j}\right), g(\varepsilon)\right]\right\}, \tag{4.54}
\end{equation*}
$$

where $h, g$ are any transformations of the random variables (including the identity transformation). ${ }^{9}$ A proof is provided in the Technical Appendix. So these estimates provide meaningful information on the original variables in the data, specifically the maximum possible strength of their correlation.

Then, eq. [4.54], combined with the fact that the Gaussian Copula nests the independence case, tells us that the estimated correlation coefficients between the transformed regressors and the error term operate also as a test for the existence of linear correlation of the untransformed variables: if these correlation measures are estimated as statistically zero, it means that the transformed variables are statistically independent, and that the untransformed variables are non-correlated linearly. While this does not preclude the existence of pure non-linear dependence between the untransformed variables, experience says that most forms of non-linear dependence manifest also as linear correlation, and so in the case the latter is zero, most likely statistical independence of the original variables can be accepted.

## Closing notes.

In this chapter we presented specifications and tools to estimate 2TSF models assuming different forms of statistical dependence. In section I, we recruited a bivariate density used in Survival \& Reliability analysis to present a specification for the 2TSF error term allowing for dependence between the two one-sided error components. To our knowledge, this is the first composite error density for the 2TSF framework that allows for such dependence.

In section II we presented a Copula model to account for regressor endogeneity. We analyzed the case of using the Gaussian Copula which has many desirable properties, and detailed each implementation step, not only for the bivariate case (which would allow to account for endogeneity of a single regressor only, since the other variable would be the error term of the regression), but also for the multivariate one. While Copula misspecification is likely to be the rule rather than the exception, we argued that we will be in a better position if we allow for endogeneity compared to ignoring it, even if we do so through a misspecified model. Also, by using the Gaussian Copula we can formally test for

[^31]misspecification and also, with the many tools available in the literature, we can assess the degree of misspecification, when it exists.

With this chapter we are done with the more technical contributions to the literature of the 2TSF framework. In the next part of the thesis, we present economic applications, where after developing models of economic phenomena and interactions, we put these tools to work.

## Chapter 5

# A targets-based Nash-bargaining 2TSF framework, with a wage determination model under productivity uncertainty. 

We develop a new 2TSF structural framework for a Nash bargaining situation that accommodates uncertainty and heterogeneous/asymmetric information, while it is based on "target prices" rather than on reservation prices. The focus is on the labor market and wage determination, but the model has much wider applicability. The model is applied to a matched employer-employee data set from Ghana. The chapter was presented in the international conference "North American Productivity Workshop NAPW-X" in June 2018, Florida USA.

In chapter 2.IV, we have discussed the Nash bargaining 2TSF framework proposed by Kumbhakar \& Parmeter (2009). The authors formulated the Nash-bargaining situation around reservation wages, extending directly a deterministic setup to a stochastic environment. The result was that their approach did not really lead to a 2TSF model, only to a single-tier SF one. To circumvent this, we develop here a framework that uses target-wages instead, and exploits the existence of asymmetric information.

## I. The Nash-bargaining equilibrium solution and its application

## in wage negotiations.

The "Nash-bargaining" model and equilibrium was introduced by Nash (1950, 1953), as a solution concept for situations where self-interested parties have an opportunity to gain through collaboration, and they must agree on how to split these gains. If they don't agree, there is no "penalty" other than the loss of the possible gain from the collaboration. The Nash-bargaining equilibrium is a solution concept in the context of Axiomatic Bargaining theory. ${ }^{1}$ It is a powerful concept because it is fully consistent with fundamental pillars of economic theory: it emerges as the unique equilibrium when the parties involved have

[^32]preferences that satisfy the basic axioms of Expected Utility theory (complete, transitive and continuous preferences, satisfaction of the Independence axiom), and it is Pareto-optimal. ${ }^{2}$

The concept was developed using a game-theoretic framework, and while it can be extended to a game with many players, by far its most frequent use is in situations with twoplayers and we will stick to that.

The setup is as follows: for a two-player game, each has some payoff if there is no agreement, denote it $s_{0}(i), i=1,2$. These "fall-back" payoffs can be strictly positive or zero, but not negative. In that way, they become "credible threats" in game-theoretic terminology: the pre-game situation of the parties won't be worsened by no-agreement, they will just lose the opportunity to gain.

Denote also $s(i), i=1,2$ the payoff of each player in case of an agreement, and denote $S_{i} \equiv s(i)-s_{0}(i)$ the "surplus function" of party $i$. Then as it has been shown (see Roth 1979), the "symmetric" Nash-bargaining solution is the argument that maximizes the product

$$
\begin{equation*}
\arg \max \left\{S_{1} \cdot S_{2}\right\}=\arg \max \left\{\left(s(1)-s_{0}(1)\right) \cdot\left(s(2)-s_{0}(2)\right)\right\}, \tag{5.1}
\end{equation*}
$$

where the decision variable is an implicit argument in the two surplus functions. It has also been shown that we can enhance the model by considering "asymmetric" situations, and still the solution

$$
\begin{equation*}
\arg \max \left\{S_{1}^{\eta} \cdot S_{2}^{1-\eta}\right\}=\arg \max \left\{\left(s(1)-s_{0}(1)\right)^{\eta} \cdot\left(s(2)-s_{0}(2)\right)^{1-\eta}\right\}, \quad 0<\eta<1, \tag{5.2}
\end{equation*}
$$

remains the unique Nash-bargaining equilibrium. The coefficient $\eta$ is frequently interpreted as reflecting inequalities in "bargaining power" or negotiating "ability". The inequalities related to $\eta$ could be written as weak, but that would give us uninteresting outcomes, where no "split of the pie" occurs.

[^33]The Nash-bargaining equilibrium concept is inherently a static-one, formulated from a technical point of view as the solution to a game in "strategic/normal form". It describes the situation before the game starts, and then it provides the solution. Binmore, Rubinstein \& Wolinsky (1986) provide insightful analysis and formulations that elaborate on the microdynamics of the negotiation process, offering a more intuitive rationale for the possible reasons why the parties do want to reach, and eventually reach, an agreement: it may be pure time-impatience, or the risk to lose the possible gain from collaboration if negotiations are prolonged, because a third party may exploit the opportunity instead. They also show that certain asymmetries can be incorporated already in the surplus functions of the players, enhancing (or restricting) the scope of interpretation of the parameter $\eta$ present in the "asymmetric" situation.

In the context of wage bargaining theory, the following mapping to the general expression [5.2] has been widely used: The players are the employee (indexed here by " $e$ ") and the firm (indexed by " $f$ "). The no-agreement payoff for the employee is her reservation wage, $s_{0}(e) \equiv \underline{\omega}$. This may reflect unemployment benefits, or another job offer that the employee has already secured. The agreement payoff of the employee is the wage $s(e) \equiv \omega$. So the surplus function for the employee is

$$
\begin{equation*}
S_{e} \equiv\left(s(e)-s_{0}(e)\right)=\omega-\underline{\omega} . \tag{5.3}
\end{equation*}
$$

The no-agreement payoff of the firm is assumed to be zero $s_{0}(f)=0 .{ }^{3}$ The agreement payoff of the firm is the value it will get from employing the worker, above the wage it will pay her. Denoting the value of the employee's output $p$ (a magnitude that is certain in a deterministic setup), we have the surplus function $s(f) \equiv p-\omega$, and since we have assumed $s_{0}(f)=0$ we get

[^34]\[

$$
\begin{equation*}
S_{f} \equiv\left(s(f)-s_{0}(f)\right)=p-\omega . \tag{5.4}
\end{equation*}
$$

\]

Note that, reasonably, $p$ is also the "maximum willingness to pay" from the part of the firm, or simply the "reservation wage" of the firm. Finally, $\eta$ is translated as the "relative bargaining power" of the employee. We assume that $\underline{\omega}<\bar{\omega}$ (which holds by construction in observed agreements, i.e. hires). Then the Nash-bargaining equilibrium wage is obtained as

$$
\begin{equation*}
\omega^{*}: \arg \max (\omega-\underline{\omega})^{\eta}(p-\omega)^{1-\eta}, \quad 0<\eta<1, \quad \omega \in[\underline{\omega}, p], p \equiv \bar{\omega} . \tag{5.5}
\end{equation*}
$$

The solution is

$$
\begin{equation*}
\omega^{*}=\eta \bar{\omega}+(1-\eta) \underline{\omega} . \tag{5.6}
\end{equation*}
$$

The higher $\eta$ is, the higher the wage will be since the solution is a convex combination and we have $\bar{\omega}>\underline{\omega}$. This validates the treatment of $\eta$ as the "relative bargaining power" of the worker side. In folklore, a negotiation is a situation where the parties engage in numerous devious tactics, manipulation of emotions, psychological warfare and other thrillers, in order to trick the other side into an agreement that is "against its own benefit" and in the benefit of the trickster. But all these are excluded in a rational framework (which moreover is realistic since we are not dealing with frivolous consumption here but with the life-supporting wages of a person and with the profits of a firm). Then, in order to determine the equilibrium/agreement point, we are only left with the actual, no-tricks "room for maneuver" each party has, inside the feasible space defined by the reservation wages. And this is accurately labeled "relative bargaining power".

The picture is one of a situation where the parties recognize that while they remain separate each in its own interests, this "game" will be successful only if there is a split in the possible gains. But this split will not be decided "collectively", under some non-economic notion of "fairness" perhaps, but it will be ultimately based on the relative bargaining power of the two parties.

We turn now to examine what modifications are needed when we introduce in the model the element of uncertainty around the value of the employee's output, as well as asymmetric information between worker and firm.

## II. Productivity uncertainty and asymmetric information: a targetsbased formulation of the Nash-bargaining wage determination game.

We consider a situation where a firm and a worker meet to examine the prospect of a match/hire of the worker in the firm (or a continuation of an existing relationship), and to negotiate over the wage of the employee. We make two realistic assumptions: first, that the value of the employee's output is no longer a certain, deterministic magnitude, neither for the firm nor for the employee (which is only reasonable since we are talking about the future). This requires that the parties form expectations about the value of this output, and expectations are based on information: our second realistic assumption is that the two parties do not have the same information sets. There are common elements, but also, each party holds private information: we have the information sets $I_{f} \neq I_{e}, I_{f} \cap I_{e} \neq \varnothing$. It follows that, except unlikely situations, we will have, adopting the conditional expectation as our chosen predictor,

$$
E\left(p \mid I_{f}\right) \neq E\left(p \mid I_{e}\right) .
$$

It may appear obvious that we just have to modify the objective function [5.5] of the deterministic setup as follows:

$$
\begin{equation*}
\omega^{*}: \arg \max (\omega-\underline{\omega})^{\eta}\left(E\left(p \mid I_{f}\right)-\omega\right)^{1-\eta}, \quad 0<\eta<1, \quad \omega \in\left[\underline{\omega}, E\left(p \mid I_{f}\right)\right], E\left(p \mid I_{f}\right) \equiv \bar{\omega}, \tag{5.7}
\end{equation*}
$$

since the presence of $p$ in the deterministic setup relates to the gains of the firm only, and it is natural to argue that the "maximum willingness to pay" for the firm in this uncertain environment is its expectation about the value of the worker's output. But the solution to [5.7] then becomes

$$
\begin{equation*}
\omega^{*}=\eta E\left(p \mid I_{f}\right)+(1-\eta) \underline{\omega}, \tag{5.8}
\end{equation*}
$$

which, from the point of view of econometric implementation suffers from two serious problems: first, over a sample of observed wage transactions, the relative bargaining power $\eta$ should realistically be treated as a variable, not as a constant, reflecting different bargaining power per worker. Then, even if we assume the usual linear regression expression for $E\left(p \mid I_{f}\right)$, we will have a specification with varying coefficients, not really estimable, except if we have panel data available and we make some rather ad hoc assumptions about $\eta$.

Second, the information set in the conditional expected value is the full information set of the firm, including private information, which is not available to the researcher as data. This intensifies the "omitted-variable" bias that is always a danger in econometric estimation.

It follows that in an environment with uncertain worker productivity, we must go beyond substituting conditional expected values for previously deterministic variables in order to arrive at an estimable specification.

Moreover, the above formulation does not take into account how the surplus functions of the two parties change during the negotiation process.

Finally, the reduced-form model in [5.8] is a single-tier SF model, which is not our target here. To obtain a 2TSF model proper, we will exploit rather than ignore the discrepancy of expectations about the worker's output, and we will also look more closely to the bargaining process itself, infusing a little bit of dynamics in our model although maintaining the static structure.

## II.1. The surplus function of the employee during the negotiation process.

Assume that we are in the process of wage bargaining, and the firm makes some specific offer to the employee, denote it $\omega_{f}^{T}$. Since we will eventually look at realized matches, we can picture a sequence of such offers converging to the equilibrium point (and converging monotonically, since in such situations concessions to the other party are almost never reversed). Since the feasible space is constrained by the initial reservation wage of the employee, $\underline{\omega}$, and so the equilibrium wage will be higher than that, it follows that eventually
some element of the sequence of the firm's offers, say the aforementioned $\omega_{f}^{T}$, is above $\underline{\omega}$, $\omega_{f}^{T}>\underline{\omega}$. If we freeze momentarily the process at this point, we realize that the options of the employee now are not "don't agree and get the fall-back payoff $\underline{\omega}$ " or "continue bargaining", but "agree now-get payoff $\omega_{f}^{T}$ " or "continue bargaining".

Since $\omega_{f}^{T}>\underline{\omega}$, the employee is now in a better position compared to before the initiation of the bargaining process. Her "credible threat" now is to accept the firm's offer, not to walk away and get her reservation wage. Therefore, her surplus function becomes

$$
\begin{equation*}
S_{e}^{\eta}=\left(\omega-\omega_{f}^{T}\right)^{\eta} . \tag{5.9}
\end{equation*}
$$

## II.2. The surplus function of the firm during the bargaining process.

Analogously, consider a sequence of counter-offers by the employee. An element of this sequence, $\omega_{e}^{T}$, will eventually be below the firm's maximum willingness to pay $E\left(p \mid I_{f}\right)$. At this point, the options of the firm are no longer "don't agree and get payoff zero" or "continue bargaining". Instead they have become "agree now - get expected payoff $E\left(p \mid I_{f}\right)-\omega_{e}^{T}>0$ " or "continue bargaining". As in the case of the employee, the firm finds itself in a better position compared to the beginning of the game, and its surplus function is now

$$
\begin{equation*}
S_{f}^{1-\eta}=\left(\left[E\left(p \mid I_{f}\right)-\omega\right]-\left[E\left(p \mid I_{f}\right)-\omega_{e}^{T}\right]\right)^{1-\eta}=\left(\omega_{e}^{T}-\omega\right)^{1-\eta} . \tag{5.10}
\end{equation*}
$$

As we have warned, while this analysis points towards a dynamic, sequential bargaining game, we want to maintain a static framework. To do this we will treat the wage offers and counter-offers not as a sequence of "tactical positions" of the parties, but as rationally formed focal targets that guide them in the negotiation process. This interpretation has weight inasmuch as every offer made has the possibility of being accepted by the other party, and so it automatically represents a credible commitment of the party that makes it. Therefore
it has to be linked to some general, pivotal strategy. We can then think of $\left(\omega_{e}^{T}, \omega_{f}^{T}\right)$ as some notion of "average" offers aligned with the goals of each respective party.

Moreover, condensing the sequential dynamic procedure into a focal point is justified up to a point by the fact that the rounds of offers and counter offers are not usually numerous. ${ }^{4}$ So we can still talk about the Nash-bargaining equilibrium as the solution to

$$
\begin{equation*}
\omega^{*}: \arg \max \left(\omega-\omega_{f}^{T}\right)^{\eta}\left(\omega_{e}^{T}-\omega\right)^{1-\eta}, \quad 0<\eta<1, \quad \omega \in\left[\underline{\omega}, E\left(p \mid I_{f}\right)\right], \quad E\left(p \mid I_{f}\right) \equiv \bar{\omega}, \tag{5.11}
\end{equation*}
$$

with solution ${ }^{5}$

$$
\begin{equation*}
\omega^{*}=\eta \omega_{e}^{T}+(1-\eta) \omega_{f}^{T} . \tag{5.12}
\end{equation*}
$$

Comparing the initial stochastic formulation (eq. [5.7]) with [5.11], we see that we have maintained the same feasible set, $\omega \in\left[\underline{\omega}, E\left(p \mid I_{f}\right)\right]$, which should be anticipated since the bargaining space has not changed. But the objective function has been transformed to reflect the improvement of the position of two parties due to the negotiation process itself, compared to their initial situation as the latter is reflected in the objective function in eq. [5.7].

What remains now is to formulate reasonable and implementable expressions for $\omega_{e}^{T}$ and $\omega_{f}^{T}$.

## II.3. The "common-information" expected output level, and the target wages

 of the employee and of the firm.Since targets are to be rationally formed, they will depend on the information possessed by each party. We argued earlier that the intersection of the two information sets $I_{f} \cap I_{e}$ is not empty: it contains characteristics (of the worker, of the firm, of the

[^35]socioeconomic environment) that are commonly accepted and used to assess/predict the output of the worker.

Then, the conditional expectation $E\left(p \mid I_{f} \cap I_{e}\right)$ is a magnitude obtained based only on shared information. It is a "symmetric-information" magnitude, an expected value based on an information set that deliberately excludes all information that would make the situation asymmetric in information. As said, it includes information from the worker's resumé, and possibly public information about the firm and the market conditions. $E\left(p \mid I_{f} \cap I_{e}\right)$ forms a common base of the two parties. But it will most likely change during the negotiation process, as information is exchanged and the set of common knowledge is altered. ${ }^{6}$ Therefore, when looking at realized matches, we distinguish between the ex ante commonknowledge information set and the ex post or equilibrium one. We assume that such disturbances of the common information set merge into a zero-mean random variable ("some news increase the expectation, some decease it"), and we write the equilibrium symmetricinformation expected output as

$$
\begin{equation*}
\mu_{p}(\mathbf{x}) \equiv E\left(p \mid I_{f} \cap I_{e}\right)+v . \tag{5.13}
\end{equation*}
$$

A useful way to think of the relation between $E\left(p \mid I_{f} \cap I_{e}\right)$ and $v$ is to consider $v$ as an enhancement/correction/adjustment of information initially included in $I_{f} \cap I_{e}$ : for example during the negotiation the worker discloses specifics or activities that were completely missing from her resumé, or elaborates further on the stated experience. Or information about the firm that the worker may initially have had is enhanced and clarified. Nevertheless, while this interpretation of the variable $v$ is plausible and theoretically sound, it is also a fragile one in empirical applications: the possible existence of omitted variables will contaminate $v$. Also, the anticipated linear formulation of $E\left(p \mid I_{f} \cap I_{e}\right)$ will be most likely an approximation, and so approximation left-overs will reside in $v$. Then we will no

[^36]longer be able to treat its effect on the outcome as strictly the effect of learning during the negotiation.

We stress that neither $E\left(p \mid I_{f} \cap I_{e}\right)$ nor $\mu_{p}(\mathbf{x})$ are the reservation wage of the worker. In general the reservation wage will be something between the "BATNA" and "WATNA" values (Best and Worst "Alternative To a Negotiated Agreement" respectively), depending on the uncertainty surrounding (or not) these alternatives, as well as the worker's risk profile. The reservation wage can be lower (or higher) than $E\left(p \mid I_{f} \cap I_{e}\right)$ or $\mu_{p}(\mathbf{x})$, depending on the specifics of the worker's situation.

We now ask, what is the target wage that the worker is after, $\omega_{e}^{T}$ ? It cannot just be $E\left(p \mid I_{f} \cap I_{e}\right)$ (except by unlikely chance) because it does not use all the information set of the worker. Moreover, we argue that individuals, being individuals, take the individualistic approach: "I am not just these characteristics. I am more than that" (and this is not necessarily some short of "psychological bias", the worker may be "right" in thinking so, based on her past performance). Moreover, beneficial particulars to the time and place can affect the worker's target wage, like for example a tip that the firm's payroll policy is "generous", or that the firm is pressed to hire fast because of rising demand.

On the other hand, detrimental specifics that could force the worker to accept a "low" wage (e.g. a pressing need for cash inflows in light of existing financial commitments) do not affect her target wage, but rather, her reservation wage. The target-wage is formed by all the things that could increase the accepted wage, not decrease it. It remains a purely desirable quantity, not something that nets positive and negative influences. We capture all these in a non-negative random variable $g \geq 0$, which we call the self-evaluation premium. Moreover, here too we are interested in the ex post target wage, the target that was when the deal was closed, and we define the worker's equilibrium target wage as

$$
\begin{equation*}
\omega_{e}^{T} \equiv \mu_{p}(\mathbf{x})+g=E\left(p \mid I_{f} \cap I_{e}\right)+v+g . \tag{5.14}
\end{equation*}
$$

Firms, on the other hand, have sometimes painful experience from how "officially able" individuals may not succeed in being productive in collaborative environments due to organizational inefficiencies, personality and cultural clashes, lack of guidance and direction: the reasons why internally produced inefficiency exists in businesses in the first place. And such situations are more frequent than pleasant efficiency surprises (because of the fundamental entropy asymmetry that makes collaboration harder than non-collaboration). So we argue that firms tend to systematically discount $E\left(p \mid I_{f} \cap I_{e}\right)$, when formulating their target wage ("it always looks better on paper than it actually is"). A firm would call that figuratively "an insurance policy", we will call it the prudential discount. As further indication of this systematic discount, we note the not-infrequent use of "fixed/variable pay" or "fixed pay plus bonus" schemes by firms, which is a way to offset the consequences of a possibly unsuccessful match while at the same time accommodating the "I am more than that" stance of the worker ("prove it and you will be rewarded"). Moreover, certain current situations (external or internal) that the firm knows of as private information may affect what the firm can expect from the match (like an imminent re-organization that will initially cost in terms of efficiency). This discount persists as the common-knowledge expected productivity is altered, because it does not depend on the specific characteristics of the worker, it is a form of "statistical bias". Analogously with the worker, here we have the equilibrium target wage of the firm

$$
\begin{equation*}
\omega_{f}^{T} \equiv \mu_{p}(\mathbf{x})-d=E\left(p \mid I_{f} \cap I_{e}\right)+v-d, \tag{5.15}
\end{equation*}
$$

where $d \geq 0$ is the "prudential discount" component. Here too $E\left(p \mid I_{f} \cap I_{e}\right)$ or $\mu_{p}(\mathbf{x})$ are not the "reservation wage" of the firm, the maximum it would be willing to pay the worker. The latter is $E\left(p \mid I_{f}\right)$, the expected value conditional on all information available to the firm, and it may in principle lie above or below $E\left(p \mid I_{f} \cap I_{e}\right)$ or eventually, $\mu_{p}(\mathbf{x})$.

And, in reverse analogy to the worker side, $d$ captures only those elements that tend to decrease the target wage of the firm, and not specifics that may force the firm to accept a higher wage. To use the same example as before, if the firm is pressed to hire fast for some
reason, this will affect the firm's "maximum willingness to pay" (while, if known to the worker, it will affect her target-wage). ${ }^{7}$

The reservation wages do not change during the negotiation. The relations between target wages and reservation wages are $\underline{\omega}<\omega_{e}^{T}, \omega_{f}^{T}<\bar{\omega}$. For a match to be feasible, we also examine only cases where $\underline{\omega}<\bar{\omega}$. But the location of the target-wage of one party with respect to the reservation wage of the other party is not constrained. For example, consider the following rather extreme situation, where targets lie outside the range defined by reservation wages:


The initial common-knowledge expected wage level $E\left(p \mid I_{f} \cap I_{e}\right)$ is located anywhere in $\left[\omega_{f}^{T}, \omega_{e}^{T}\right]$, even outside the range defined by reservation wages. But the feasible space for the equilibrium wage is always the interval $[\underline{\omega}, \bar{\omega}]$.

## II.4. The Nash bargaining solution in the targets-based model.

Inserting the expressions for the target wages [5.14] and [5.15] into the objective function [5.12] we have the Nash-bargaining solution

$$
\begin{align*}
& \omega^{*}=\eta\left[E\left(p \mid I_{f} \cap I_{e}\right)+v+g\right]+(1-\eta)\left[E\left(p \mid I_{f} \cap I_{e}\right)+v-d\right] \\
& \Rightarrow \omega^{*}=E\left(p \mid I_{f} \cap I_{e}\right)+v+\eta g-(1-\eta) d \tag{5.16}
\end{align*}
$$

The unobservable composite term $v+\eta g-(1-\eta) d$, has the 2TSF composite error form.

[^37]The systematic observable component $E\left(p \mid I_{f} \cap I_{e}\right)$ is the ex ante, the initial symmetric-information part of the wage (and certainly not the full-information wage, more on that in a while). The term $v$ captures shifts in the common-knowledge expected productivity during negotiations, and other unobserved elements. The term tending to increase the equilibrium wage, $\eta g$, is the self-evaluation premium of the worker weighted by the worker's bargaining power. That product can be labeled the "bargaining performance" of the worker. The term tending to decrease the wage (due to the negative sign) $(1-\eta) d$ is the prudential discount by the firm, weighted by its own relative bargaining power. Correspondingly, this term can be labeled the "bargaining performance" of the firm.

Equation [5.16] represents a rich combination of actual aspects of economic transactions with bargaining: there is a common ground of rational thinking that both parties accept, $E\left(p \mid I_{f} \cap I_{e}\right)$; there is their "subjective views of the world" through the terms $g$ and $d$; and there is the allocation of bargaining power between them, $(\eta, 1-\eta)$ that reflects their constraints under which they operate, constraints that force them to accept an outcome less than "their wish" (more on that in a while). Already as a qualitative interpretational tool, eq. [5.16] explains well different outcomes in various labor markets. Consider for example unskilled labors: Common-information expected productivity $E\left(p \mid I_{f} \cap I_{e}\right)$ will be relatively low, and workers aspirations $g$ also. In an industry where unskilled labor is a peripheral input of production (say, clerical jobs in the IT industry), relative bargaining power of the worker $\eta$ will also be low, and all these together conspire towards the low wages observed. But if the sector is in farming, say fruit-picking, unskilled labor is a core input, and the relative bargaining power of the worker increases -hence the comparatively higher wages observed.

Also [5.16] reflects a degree of dynamics: $E\left(p \mid I_{f} \cap I_{e}\right)$ will provide information on the situation at the beginning of the negotiation process, while $\omega^{*}-E\left(p \mid I_{f} \cap I_{e}\right)=\varepsilon=v+w-u$, i.e. the residual, is the quantitative result of the negotiation itself, the "net negotiation effect".

[^38]

Note that the equilibrium wage will not in general equal the equilibrium commonknowledge expected productivity $E\left(p \mid I_{f} \cap I_{e}\right)+v$, because the one-sided components will not, in general, be equal. And the higher the self-evaluation premium of the employee the higher the agreed wage, while the higher the prudential discount by the firm the lower the agreed wage. This conforms with the real-world observation that such "subjective" elements (but not hand-waiving tricks) do matter in an economic negotiation, in addition to the allocation of bargaining power, which has a more "objective" flavor: in fact eq. [5.16] tells us that, regarding the worker for example, a low bargaining power (low $\eta$ ) can be mitigated by a higher self-premium (higher $g$ ) from the part of the worker, which we can picture as a situation where the worker actively promotes herself in the negotiations, showing selfconfidence and belief in her abilities, productivity and prospects with the firm, even though she is, say, a young worker with little to show in terms of accomplishments.

It follows that any deviation of the equilibrium wage above or below $E\left(p \mid I_{f} \cap I_{e}\right)$ is not just "surplus extraction due to bargaining power": it is a mixed effect of the allocation of bargaining power and of how the parties conduct themselves in the negotiations, and this is why we believe that "bargaining performance" is an appropriate label for it.

Finally, as already noted, the existence of an agreed wage does not imply that at the time of agreement, targets coincide with it: I can very consciously agree to a wage level because it is above my reservation wage, even though I would want for the agreement to be more in my favor. Still I agree, because I recognize the limits of my bargaining power and any other constraints under which I operate. ${ }^{9}$ So we do not argue that the observed wage is an equilibrium in Edmund Phelps' sense of "equilibrium as fulfilled expectations". It may be, but it may be not: in this last case the agreement may be "temporary" in the eyes of the parties, although they will certainly keep this to themselves as another piece of private information. A worker may agree to the wage and start working with the firm while she keeps looking for another job, and the firm may agree to the wage and employ the worker, while keep looking for a replacement. The rationalization here is that further delay of actual

[^39]productive work may hurt both or at least one of the two parties, and so they concede temporarily while they keep searching for a better long(er)-term deal elsewhere.

## II.5. Market-level implications, strong predictions, and logarithmic transformations.

If we apply the unconditional expected value operator to eq. [5.16] we get

$$
\begin{align*}
& E\left(\omega^{*}\right)=E\left[E\left(p \mid I_{f} \cap I_{e}\right)\right]+E(v)+E(\eta g)-E[(1-\eta) d] \\
& \Rightarrow E\left(\omega^{*}\right)=E(p)+E(w)-E(u), \tag{5.17}
\end{align*}
$$

where we applied the law of iterated expectations, and we have used the usual symbols for the one-sided error components $w \equiv \eta g, u \equiv(1-\eta) d$. Since $\omega^{*}$ is the observed wage, and we will obtain estimates for $E(w), E(u)$, we can obtain also an estimate for the unconditional expected value of the future output of the workers through

$$
\begin{equation*}
\hat{E}(p)=\frac{1}{n} \sum_{i=1}^{n} \omega_{i}^{*}-\hat{E}(w)+\hat{E}(u) . \tag{5.18}
\end{equation*}
$$

Another market-level phenomenon we may study is the following: our model incorporates the assumption that $\omega_{i}^{*} \leq E\left(p_{i} \mid I_{f, i}\right) \forall i$, because we have argued that $E\left(p_{i} \mid I_{f, i}\right)$ is the maximum willingness to pay from the part of a firm. But this inequality over all transactions implies also that

$$
\begin{equation*}
E\left(\omega_{i}^{*}\right) \leq E\left[E\left(p_{i} \mid I_{f, i}\right)\right]=E\left(p_{i}\right) \tag{5.19}
\end{equation*}
$$

Combining [5.17] and [5.19] we have

$$
\begin{equation*}
E(p)+E(w)-E(u) \leq E(p) \Rightarrow E(w) \leq E(u) . \tag{5.20}
\end{equation*}
$$

This is a strong and testable prediction of the model. It says that at the sample level we should obtain $\hat{E}(w) \leq \hat{E}(u)$, so that the theoretical model is not contested by the data. This result has a strong interpretation in economic terms: it says that on average, the self evaluation premium of the workers should always be lower than the prudential discount of the firms (as weighted by the relative bargaining power), irrespective of any other factors at play. Since the model we have built is in principle applicable to all labor markets (and beyond), this appears to be a worryingly universal result. But, thankfully, it stems directly from a specific assumption of the theoretical model, and so it is not really a universal conclusion but a way to test this specific assumption.

When building the Nash-bargaining environment, we assumed that the "noagreement" payoff of the firm was zero. It is under this assumption that $E\left(p_{i} \mid I_{f, i}\right)$ becomes the maximum willingness to pay form the part of the firm and so we obtain the inequality $\omega_{i}^{*} \leq E\left(p_{i} \mid I_{f, i}\right)$ and the prediction $E(w) \leq E(u)$.

Allow now for a more general formulation where the no-agreement payoff for the firm, say $p_{f}^{(n a)}$, can be positive or negative, or even zero (and unknown to the researcher). It could be positive if the firm has tentatively agreed with some other prospective worker. It could be negative if there is no alternative and the firm stands to lose if it doesn't hire the worker, say, because there is a client's contract that will remain unfulfilled and penalty clauses will be activated. We stress that "loss" here means tangible dead-burden costs incurred, and not "non-realized profit opportunities", because profit prospects are already incorporated in $E\left(p_{i} \mid I_{f, i}\right)$. In such a situation, the solution of the model is not affected, but the maximum willingness to pay from the part of the firm becomes now $E\left(p_{i} \mid I_{f, i}\right)-p_{f, i}^{(n a)}$ (it is subtracted, because we mapped a positive $p_{f, i}^{(n a)}$ to something beneficial for the firm). Then the model predicts

$$
\begin{equation*}
E(w)-E(u) \leq-E\left(p_{f}^{(n a)}\right) \Rightarrow E\left(p_{f}^{(n a)}\right) \leq E(u)-E(w) . \tag{5.21}
\end{equation*}
$$

We see that if we obtain $\hat{E}(w)<\hat{E}(u)$ then $E\left(p_{f}^{(n a)}\right)$ can be positive, zero or negative, with $\hat{E}(u)-\hat{E}(w)$ being an upper bound for the average positive no-agreement payoff for the firms. But if we obtain $\hat{E}(w)>\hat{E}(u)$ we have market-level evidence that $E\left(p_{f}^{(n a)}\right)<0$ and that on average the firms in the market are really pressed to hire. The result is that here Labor on average earns more that the value of its output, due to the existence of deadburden costs for the firms if they don't hire. Also, this provides indirect information on the relative bargaining power $\eta$, which is consistent with the model from another angle since $w \equiv \eta g, u \equiv(1-\eta) d$.

## Logarithmic transformations.

The previous discussion was implicitly based on a regression model where the dependent variable, the wage, enters as is, in its level. What happens to these two marketlevel conclusions (the estimation of the average value of a worker's output, and relative bargaining power), when, as is often the case, we apply a logarithmic transformation to the dependent variable?

Fundamentally, the relation we examine is between the wage and the conditional expected value of the worker's future output, and not between the wage and the worker's future output per se. So we relate $\omega_{i}^{*} \rightarrow E\left(p \mid I_{f} \cap I_{e}\right), \omega_{i}^{*} \rightarrow E\left(p_{i} \mid I_{f, i}\right)$, and therefore, going logarithmic leads to the relationships

$$
\ln \omega_{i}^{*} \rightarrow \ln \left[E\left(p \mid I_{f} \cap I_{e}\right)\right], \ln \omega_{i}^{*} \rightarrow \ln \left[E\left(p_{i} \mid I_{f, i}\right)\right] .
$$

As regards the average value of a worker's output, if we stay at the logarithmic level and take the expected value, we will obtain

$$
E\left(\ln \omega_{i}^{*}\right)=E\left\{\ln \left[E\left(p \mid I_{f} \cap I_{e}\right)\right]\right\}+E(w)-E(u),
$$

and the law of iterated expectations cannot be applied. Moreover, because the logarithm is a concave function we have by Jensen's inequality

$$
E\left\{\ln \left[E\left(p \mid I_{f} \cap I_{e}\right)\right]\right\}<\ln \left\{E\left[E\left(p \mid I_{f} \cap I_{e}\right)\right]\right\}=\ln [E(p)]
$$

so what we obtain here is

$$
E\left(\ln \omega_{i}^{*}\right)<\ln [E(p)]+E(w)-E(u) \Rightarrow E(p)>\exp \left\{E\left(\ln \omega_{i}^{*}\right)-E(w)+E(u)\right\},
$$

i.e. only a lower bound for $E(p)$. If we first go back in levels, and then apply the expected value, we have

$$
\omega_{i}^{*}=E\left(p \mid I_{f} \cap I_{e}\right) \cdot \exp \{v+w-u\} \Rightarrow E\left(\omega_{i}^{*}\right)=E\left[E\left(p \mid I_{f} \cap I_{e}\right) \cdot \exp \{v+w-u\}\right]
$$

As we will see in a while, given our assumptions we can separate the component $\exp \{w-u\}$ but not $\exp \{v\}$. So we will arrive at

$$
E\left(\omega_{i}^{*}\right)=E\left[E\left(p \mid I_{f} \cap I_{e}\right) \cdot \exp \{v\}\right] \cdot E[\exp \{w-u\}],
$$

and again we cannot apply the law of iterated expectations to obtain $E(p)$.

Turning to the issue of relative bargaining power and hiring pressure, we can define the no-agreement payoff for the firm multiplicatively, $E\left(p_{i} \mid I_{f, i}\right) \cdot p_{f}^{(n a)}$, where here $p_{f}^{(n a)}$ is smaller, equal or higher than unity. Then

$$
\omega_{i}^{*} \leq E\left(p_{i} \mid I_{f, i}\right) \cdot p_{f}^{(n a)} \Rightarrow E\left[\omega_{i}^{*}\right] \leq E\left[E\left(p_{i} \mid I_{f, i}\right) \cdot p_{f}^{(n a)}\right]=E\left[E\left(p_{i} \mid I_{f, i}\right)\right] \cdot E\left[p_{f}^{(n a)}\right]
$$

$$
\Rightarrow E\left(\omega_{i}^{*}\right) \leq E(p) \cdot E\left[p_{f}^{(n a)}\right] .
$$

We have separated the expected value of the product because reasonably the noagreement payoff for the firm is independent of the expected value of the worker's output. But the left hand side will again be as previously, and we won't be able to obtain $E(p)$ also in the left hand side, then simplify and so obtain a relation between, $w, u, p_{f}^{(n a)}$ only. Going at the logarithmic level, and then apply the expected value, produces the same inconclusive results.

So the logarithmic regression setup comes at a price.

## II.6. Contrast with the Polacheck \& Yoon (1987) 2TSF framework.

Finally, we have to contrast, interpretation-wise, our model, eq. [5.16] here, with the foundational 2TSF structural framework, that of Polacheck \& Yoon (1987), see chapter 2.I, eq. [2.13] there:

Table 1. Comparison of 2TSF structural frameworks for the Labor market.

| Specification component | Polachek \& Yoon (1987) 2TSF framework | 2TSF targets-based Nashbargaining framework |
| :---: | :---: | :---: |
| Systematic part | $\omega_{F I}+\left(\mathbf{x}_{i}-\mathbf{x}_{F I}^{R}\right)^{\prime} \boldsymbol{\beta}$ <br> Full-information wage plus deviations due to characteristics deviating from the "representative" firm-worker pair. | $E\left(p \mid I_{f} \cap I_{e}\right)$ <br> Common-information expected output. |
| Zero-mean disturbance | Linear approximation left-overs and omitted variables. | Learning effects during the negotiation, linear approximation left-overs and omitted variables. |
| One-sided error components | Effects of incomplete and private information. | Effects of bargaining power and of private information. |

We see that the two are not really in conflict: "full-information wage plus deviations due to individual firm/worker characteristics", is not incompatible with "expected productivity based on common information" -in fact the latter can be seen as a version of the
former with the uncertainty of the situation brought in the forefront. As regards the components of the composite error term, they pretty much represent similar things. For example the fact that in Polachek \& Yoon (1987) the issue of bargaining power may not be explicitly mentioned, but enhancing the role of the one-sided components is not contradicting their previous narrower interpretation.

## II.7. Buts and rebuttals.

We can think of three objections to the 2TSF Nash-bargaining framework we have built.

The first is that, since each party formulates a target, shouldn't its attempt be to minimize the distance of the agreed wage from its own target? While this sounds intuitive, it suffers from a subtle but critical flaw: it essentially elevates the targets as being the "absolute best" outcomes for each party. This would be misleading: if for some reason during the negotiation the one party is offered something better than its own target, the party will certainly take it. Targets are benchmarks, guiding principles and negotiation drivers, not ultimate goals, not even thresholds like reservation wages are. And moreover, we showed clearly how the target of one party becomes the running "reservation wage" of the other party, as it materializes into an offer. This makes the correct goal here to be "maximize the distance of the wage from the target of the other party" and not "minimize the distance of the wage from one's own target".

The second objection could be that by not placing targets at the extremes, we violate basic postulates related to the maximizing behavior of the firm and of the worker: the target wage of the firm should invariantly be "zero", and the target wage of the worker should be "the whole value of her output", this argument would go. While this could conceivably happen, it is not realistic as a model of behavior observed most often, a model of how market participants negotiate in the majority of the cases. It is well known that a negotiation can prematurely break down if one of the parties appears "unreasonable" in its demands. Parties try to form realistic expectations and targets given the specifics and the constraints of the situation. Going back to the diagrams with the allocation of target-wages relative to the reservation wages, consider again the situation


Here, the parties adopt a "tough" negotiating stance. This can be deliberate, in order to structure the expectations of the other party as to what it could "reasonably" expect from the situation. But it also runs the risk of signaling that a deal is not possible. It is more likely that the actual situation will look like


This still reflects each party's attempt to pursue its own interests and influence the expectations of the other party, but it allows the two sides to "stay connected" during negotiations, and we assume that they came to the negotiating table in order to strike a deal, and not to pass the time enjoying the pleasures of negotiating while a priori expecting the negotiation to fail. Also, we have stressed the fact that the parties here do not try to achieve their target and stay there.

We could perhaps further rationalize such "equilibrium seeking" behavior if we invoked Herbert A. Simon's "satisficing" behavioral postulate, but we don't, remaining strictly into a full-rationality framework (under incomplete information of course). And we should remember that we want a model in order to study the negotiations that concluded in an agreement and not those that broke down.

The third objection could point out that the initial reservation wages do exist and they do not appear to control the bargaining procedure in this model. But since we will implement the model on realized matches, it follows that the observed/equilibrium wage did not violate the thresholds imposed by the reservation wages, even though we did not deal with them explicitly in our formulation. Initial reservation wages do affect the whole process in the appropriate static way: by determining the boundaries of the negotiation space.

## II.8. Applicability in other bargaining situations.

In order for the 2TSF Nash-bargaining framework to be applicable, the real-world phenomenon under study must have the following properties:

1) Asymmetric and heterogeneous information
2) Target-formulation from the negotiating parties
3) Systematic mark-up from the seller and systematic discount from the buyer, of the symmetric-information expected value.
4) Uncertainty about the total gain over which bargaining takes place.

Properties $1 \& 2$ exist in almost all cases of bargaining. Property 3 is also observed, either in the "honest" variant described earlier, or just as a bargaining tactic (which links the Axiomatic and Strategic Bargaining approaches). Such tactics are the oldest trick in the book, almost an obligatory ritual in flea-market transactions, and still widely used around the world in all sorts of trades. ${ }^{10}$ Property 4 is more subtle but it is expected to hold. It can reflect genuine uncertainty linked to the future, but it may also reflect diverging assessments of this value due to informational heterogeneity, free of the time-aspect. So the applicability of the model becomes almost universal, not just as a measuring but also as an interpretational tool.

As examples, we turn to look at papers where the a 2TSF bargaining framework of has been invoked and applied.

Kinukawa \& Motohashi $(2010,2016)$ used the model to examine alliances in the biotechnology market. Here, sellers are biotechnology companies, and buyers usually are pharmaceuticals. This market is characterized by an above-average level of uncertainty. So we expect that while sellers focus on the prospects for a medical/commercial breakthrough, the monopolistic rents that the buyer of their inventions will enjoy etc, and so will tend to value their new biotechnologies above the symmetric-information expected value, pharmaceuticals will not be able to forget all the R\&D that never led to marketable drugs, for any number of reasons, and so will tend to discount the pivot of the negotiation. Moreover, uncertainty about the value of what is under negotiation certainly exists.

[^40]Wang Y (2016a, 2016b) applied it to foreign aid in an aid-for-policy bargaining framework. Here the sellers are the recipient countries, and the buyers are the donors. What is sold and so under evaluation, is the actual power/willingness of sellers to implement policies requested by the donors, something that evidently is uncertain. We also expect large informational asymmetries and so that the politicians on both sides will tend to mark-up or discount the symmetric-information expected value of what is sold here, due to large private information sets, and what this entails for the other side.

Zhang, Zhang, Yang \& Zhou (2017) applied it to tourist shopping, considering tourist buyers and local sellers of a product or service. The situation is also characterized by big informational asymmetries. Regarding the uncertainty surrounding the total gain here, it is easy to picture it from the point of view of the buyer: the caveat emptor dictum, or, closer to home, the notorious "lemon market" for used cars, express exactly the uncertainty that a buyer faces. But why a seller of a tangible good, or of a short-term service may be uncertain about the total gain? Here, this aspect of uncertainty is better thought of as "uncertainty about what value does the good/service has for the buyer".

## II.9. Econometric specification, endogeneity and dependence.

We re-focus on the equilibrium wage equation. We have arrived at the 2TSF specification

$$
\begin{gather*}
\omega_{i}=E\left(p_{i} \mid I_{f} \cap I_{e}\right)+v_{i}+w_{i}-u_{i}, \quad i=1, \ldots, n \\
v_{i} \sim N\left(0, \sigma_{v}^{2}\right), \quad w_{i}=\eta_{i} g_{i}, \quad u_{i}=\left(1-\eta_{i}\right) d_{i}, \tag{5.22}
\end{gather*}
$$

where we have included the observation index on purpose, and we have assumed normality of the $v_{i}$ term, as is customary. We explicitly model the relative bargaining power of the workers $\eta_{i}$ as a random variable, whose realization differs from worker to worker.

The first thing to stress here is that the way the conditional expected value is defined, it requires to have available regressors that characterize both sides of the transaction. For the equilibrium wage equation this means the use of a matched employer-employee data set. If we attempt to specify the standard Mincerian "earnings" equation, where usually data
pertaining only to the worker are used, we will not really be estimating $E\left(p_{i} \mid I_{f} \cap I_{e}\right)$, but some other conditional expected value. Note also that the symmetric information set here is common to all observations (as regards the kind of information it includes, not the actual values per observation).

As regards the issue of endogeneity of the regressors that will be used to model the observable systematic part $E\left(p \mid I_{f} \cap I_{e}\right)$, we note that the terms $w_{i}, u_{i}$ represent elements that are functions of all other information than $I_{f} \cap I_{e}$. So by construction, the regressors will be exogenous to these components. But $v_{i}$ represents changes in the information set $I_{f} \cap I_{e}$ during the negotiations. Therefore it may be the case that the regressors will be correlated with the symmetric component of the composite error term. Moreover, the realities of actual empirical research can induce endogeneity due to correlation, if a regressor known to usually be "common knowledge" in the transaction under study is not available to the researcher as data, and it is thought to be correlated with the available regressors. We can deal with this possible endogeneity issue by applying the Copula model presented in chapter 4.

As regards dependence between the one-sided error components, by allowing relative bargaining power to vary for each worker, we obtain correlation between the one sided terms $w, u$. We do assume that the random variable $\eta_{i}$ is independent of the other elements and we have, using $\mu$ to denote the mean,

$$
\begin{aligned}
\operatorname{cov}\left(w_{i}, u_{i}\right) & =E\left[\eta_{i} g_{i}\left(1-\eta_{i}\right) d_{i}\right]-E\left[\eta_{i} g_{i}\right] E\left[\left(1-\eta_{i}\right) d_{i}\right] \\
& =E\left[\eta_{i}\left(1-\eta_{i}\right)\right] \mu_{g} \mu_{d}-E\left(\eta_{i}\right) E\left(1-\eta_{i}\right) \mu_{g} \mu_{d}=\operatorname{cov}\left(\eta_{i}, 1-\eta_{i}\right) \mu_{g} \mu_{d},
\end{aligned}
$$

while $\operatorname{cov}\left(\eta_{i}, 1-\eta_{i}\right)=E\left(\eta_{i}-\eta_{i}^{2}\right)-E\left(\eta_{i}\right) E\left(1-\eta_{i}\right)=-\sigma_{\eta}^{2}<0$.
The model predicts negative correlation between the two components. Therefore, in an econometric application we should use a 2 TSF composite error density that incorporates dependence between the two one-sided components, like the Correlated Exponential one we have presented in chapter 4.

## III. An empirical application.

We will implement the model using part of a matched employer-employee data set that has been originally used in Bigsten et al. (2000). ${ }^{11}$ The purpose of their study was to examine the rates of return on physical and human capital in Africa's manufacturing sector. The data were collected in three survey-waves, they refer to the early 90 's and come from five African countries.

We will use the data related to Ghana that refer to the years 1992-1994. The full sample size was 2,975 observations, but monthly employee earnings were recorded only for 2,565 of them. Another 8 observations were dropped as outliers with unrealistically low reported monthly earnings, and so Bigsten et al. (2000) used a sample of size 2,557. In our case and in order to include all the co-variates deemed relevant to our purposes, the final sample size was further reduced to $n=1,910$. The composition of the sample we used as regards number of employees and firms per survey wave is as follows:

Table 2. Composition of the Ghana sample

|  | No of firms | No of employees <br> (observations) |
| :--- | ---: | ---: |
| Wave 1 (1992) | 110 | 474 |
| Wave 2 (1993) | 119 | 562 |
| Wave 3 (1994) | 121 | 874 |
|  | Sample size $=$ | $\mathbf{1 9 1 0}$ |

Firm size as indicated by number of employees varied greatly, from 2 to $\sim 600$ people.
We pooled the data and used a linear trend $t=0,1,2$ to represent the possible timeeffect. The dependent variable is the log of monthly earnings (basic wages plus allowances) expressed in US PPP dollars as calculated in the original study. So this is a semi-log

[^41]specification. Descriptive statistics of the dependent variable and the explanatory variables are presented in the next table:

Table 3. Descriptive Statistics

| Type | Label | Description | Mean | S.D. | Min | Max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Dep var | LMEARNPP | log monthly earnings in US PPP dollars | 4.68 | 1.02 | 1.68 | 7.55 |
| Employee characteristics | MALE <br> TENURE <br> EDUC <br> PEXP <br> MGMT <br> SUPER <br> ADMIN <br> SALES <br> TECH | Binary Sex dummy (1= male) <br> Tenure in the firm in years <br> Formal education in years <br> Potential Experience (maximum possible) <br> Binary hierarchical dummy (1=management) <br> Binary hierarchical dummy (1=supervisor) <br> Binary functional dummy ( $1=$ administration) <br> Binary functional dummy ( $1=$ sales) <br> Binary functional dummy (1=technician) | $\begin{array}{r} 0.83 \\ 5.72 \\ 10.54 \\ 14.52 \\ 0.04 \\ 0.07 \\ 0.08 \\ 0.06 \\ 0.06 \end{array}$ | $\begin{array}{r} 0.38 \\ 6.45 \\ 3.49 \\ 10.86 \\ 0.20 \\ 0.26 \\ 0.26 \\ 0.24 \\ 0.24 \end{array}$ | $\begin{array}{r} 0.00 \\ 0.01 \\ 0.00 \\ -5.00 \end{array}$ | $\begin{array}{r} 1.00 \\ 48.00 \\ 19.00 \\ 59.00 \end{array}$ |
| Firm characteristics | CAPCITY <br> WOOD1 <br> TEXTILE1 <br> METAL1 <br> ANYFOR <br> ANYSTAT <br> LEMP <br> LVADEMPPPP <br> LCAPEMPPPP | Socioeconomic dummy ( $1=$ firm in country's capital) Industry Sector (1= wood sector) <br> Industry Sector ( $1=$ textile and clothing sector) Industry Sector (1= metal and machinery sector) Ownership dummy ( $1=$ some foreign equity) Ownership dummy (1= some State equity) log of number of employees log of Value-added/employee in US PPP dollars log of Physical capital /employee in US PPP dollars | $\begin{aligned} & 0.61 \\ & 0.33 \\ & 0.02 \\ & 0.31 \\ & 0.24 \\ & 0.08 \\ & 3.59 \\ & 8.06 \\ & 7.69 \end{aligned}$ | $\begin{aligned} & 0.49 \\ & 0.47 \\ & 0.15 \\ & 0.46 \\ & 0.43 \\ & 0.28 \\ & 1.14 \\ & 1.24 \\ & 2.11 \end{aligned}$ | $0.69$ $1.51$ $2.51$ | $6.43$ $10.86$ $11.64$ |

Except of the above co-variates, the specification included also the squares of the variables Tenure, Education and Potential Experience, a constant term, and the time trend mentioned earlier.

We will treat as potentially endogenous two variables: Education, to acknowledge its possible correlation with the unobservable "ability", and "Potential Experience", since it is by design an inaccurate (specifically, maximal) proxy for actual experience, and so the related measurement error resides in the error term.

We have the wage-bargaining reduced form equation

$$
\omega_{i}^{*}=E\left(p_{i} \mid I_{f} \cap I_{e}\right)+v_{i}+\eta_{i} g_{i}-\left(1-\eta_{i}\right) d_{i},
$$

which in generic notation becomes $y_{i}=\mathbf{x}_{i}^{\prime} \beta+\varepsilon_{i}$ with the mapping

$$
\left\{\begin{array}{cc}
\omega_{i}^{*}=y_{i} & \varepsilon_{i}=v_{i}+w_{i}-u_{i} \\
E\left(p_{i} \mid I_{f} \cap I_{e}\right)=\mathbf{x}_{i}^{\prime} \beta & , \quad w_{i}=\eta_{i} g_{i} \\
u_{i}=\left(1-\eta_{i}\right) d_{i}
\end{array} .\right.
$$

The econometric specification should take into account both the dependence between the two one-sided error components, as well as the possible dependence of the regressors with the error term. These will be modeled by the tools we developed in chapter 4, the first by the Correlated Exponential 2TSF specification and the second by the Gaussian copula.

Regarding the Copula model, we must start by considering the joint density of the endogenous regressors, say $\widehat{\mathbf{x}}$, and of the error term

$$
f(\widehat{\mathbf{x}}, \varepsilon)=f(\widehat{\mathbf{x}}) \cdot f_{\varepsilon}(\varepsilon) \cdot c\left(F(\widehat{\mathbf{x}}), F_{\varepsilon}(\varepsilon)\right),
$$

where $c\left(F(\widehat{\mathbf{x}}), F_{\varepsilon}(\varepsilon)\right)$ is the copula density of interest. Given that we have two potentially endogenous variables, the Gaussian copula density will include three variables, the two transformed regressors and the distribution function of the composite error term. Omitting the terms that do not contain parameters of interest, the log-likelihood is, in generic notation

$$
\left.\begin{array}{rl}
\ln \breve{L}_{i}=-\frac{n}{2} \ln \operatorname{det}(\breve{\mathrm{R}})-\frac{1}{2} \sum_{i=1}^{n} \mathbf{q}_{i}^{\prime} \breve{\mathrm{R}}^{-1} \mathbf{q}_{i}+ & \frac{1}{2} \sum_{i=1}^{n} \tag{5.23}
\end{array} \Phi^{-1}\left(F_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right)\right)\right]^{2}, ~+\sum_{i=1}^{n} \ln f_{\varepsilon}\left(y_{i}-\mathbf{x}_{i}^{\prime} \beta ; \boldsymbol{\theta}\right),
$$

where $\Phi$ represents the standard Normal distribution function. The vector $\mathbf{q}_{i}$ contains the two transformed stochastic regressors and the distribution function of the composite error term, and $\breve{\mathrm{R}}$ is their correlation matrix, here

$$
\breve{\mathrm{R}}=\left[\begin{array}{ccc}
1 & \hat{\rho}_{12} & \rho_{1 \varepsilon} \\
\hat{\rho}_{12} & 1 & \rho_{2 \varepsilon} \\
\rho_{1 \varepsilon} & \rho_{2 \varepsilon} & 1
\end{array}\right]
$$

where $\hat{\rho}_{12}$ will be calculated as the sample correlations of the two stochastic regressors, e.g. $\hat{\rho}_{12}=\operatorname{corr}\left(\Phi^{-1}\left(\breve{x}_{1}\right), \Phi^{-1}\left(\breve{x}_{2}\right)\right)$ with $\breve{x}_{i}=\hat{F}_{i}\left(x_{i}\right)$, and the MLE will estimate $\breve{\boldsymbol{\rho}}_{x \varepsilon}=\left(\rho_{1 \varepsilon}, \rho_{2 \varepsilon}\right)$.

For the error term we will assume the Correlated Exponential 2TSF distribution, where the symmetric error component follows a $N\left(0, \sigma_{v}^{2}\right)$ distribution while the two one-sided components follow jointly Freund's Bivariate Exponential extension. The density and distribution functions of the composite error term are

$$
\begin{align*}
& f_{\varepsilon}(\varepsilon)=\exp \left\{-\frac{1}{2}\left(\varepsilon / \sigma_{v}\right)^{2}\right\}\left[m b^{\prime} \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)+(1-m) a^{\prime} \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)\right]  \tag{5.24}\\
& F_{\varepsilon}(\varepsilon)=\Phi\left(\frac{\varepsilon}{\sigma_{v}}\right)+\exp \left\{-\frac{1}{2}\left(\varepsilon / \sigma_{v}\right)^{2}\right\}\left[m \exp \left\{\frac{1}{2} \omega_{2}^{2}\right\} \Phi\left(-\omega_{2}\right)-(1-m) \exp \left\{\frac{1}{2} \omega_{3}^{2}\right\} \Phi\left(\omega_{3}\right)\right] \tag{5.25}
\end{align*}
$$

$$
\text { with } \omega_{2} \equiv \frac{\varepsilon}{\sigma_{v}}+b^{\prime} \sigma_{v}, \quad \omega_{3} \equiv \frac{\varepsilon}{\sigma_{v}}-a^{\prime} \sigma_{v} \quad \text { and } \quad \sigma_{v}, a^{\prime}, b^{\prime}>0, \quad 0<m<1
$$

At the preparation stage for the Copula, the Education variable required first the "continuation transformation" in order for its empirical probabilities to be distributed as Uniform $(0,1)$ (see chapter 4).

We have shown in chapter 4 that the Gaussian Copula is the correct specification if the transformed variables included in it follow jointly a multivariate Normal distribution (MVN). We performed three such MVN tests for the transformed variables of Education and Potential Experience. ${ }^{12}$ The results were mixed: at the $5 \%$ significance level, Mardia's

[^42]skewness-kurtosis test rejected multivariate normality on account of skewness but not on account of kurtosis. Henze \& Zinkler's test rejected MVN, while Royston's test did not reject it. As a visual aid, the chi-square Q-Q plot shows that the specification can be tolerated:

Figure 1: MVN Chi-square Q-Q plot for Education and Potential Experience.


For purposes of contrast and comparison, we estimated four different models: Ordinary Least Squares, the benchmark Exponential 2TSF specification that assumes joint independence of the error component and exogenous regressors, the Correlated Exponential 2TSF specification assuming exogenous regressors, and this last one with a copula density attached. Table 4a contains the estimates for the regression coefficients.

Table 4a: Estimates of regression coefficients.

| Regressor | OLS | Ind. Exp. 2TSF | Corr. Exp. 2TSF | Corr. Exp. 2TSF with Copula |
| :---: | :---: | :---: | :---: | :---: |
| const | 1.5612 | 1.803 | 1.6933 | 3.5543 |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
| MALE | -0.0405 | 0.0122 | 0.0120 | 0.1284 |
|  | (0.356) | (0.763) | (0.768) | (0.000) |
| TENURE | 0.0139 | 0.0123 | 0.0123 | 0.0137 |
|  | (0.045) | (0.047) | (0.048) | (0.009) |
| TENSQ | -0.0003 | -0.0002 | -0.0002 | -0.0002 |
|  | (0.270) | (0.409) | (0.409) | (0.252) |
| EDUC | -0.008 | -0.028 | -0.028 | -0.0221 |
|  | (0.511) | (0.021) | (0.020) | (0.038) |
| EDUCSQ | 0.0039 | 0.0045 | 0.0045 | 0.0017 |
|  | (0.000) | (0.000) | (0.000) | (0.004) |
| PEXP | 0.0787 | 0.0694 | 0.0694 | -0.0078 |
|  | (0.000) | (0.000) | (0.000) | (0.154) |
| PEXPSQ | -0.0013 | -0.0012 | -0.0012 | -0.0003 |
|  | (0.000) | (0.000) | (0.000) | (0.005) |
| MGMT | 0.7662 | 0.7594 | 0.7591 | 0.7856 |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
| SUPER | 0.3623 | 0.3619 | 0.3619 | 0.3284 |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
| ADMIN | 0.2916 | 0.2708 | 0.2709 | 0.2475 |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
| SALES | 0.2270 | 0.2117 | 0.2115 | 0.2199 |
|  | (0.001) | (0.001) | (0.001) | (0.000) |
| TECH | 0.2418 | 0.2064 | 0.2062 | 0.1232 |
|  | (0.000) | (0.000) | (0.000) | (0.010) |
| CAPCITY | 0.1786 | 0.1805 | 0.1806 | 0.1591 |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
| W00D1 | 0.0197 | 0.0785 | 0.0789 | 0.1078 |
|  | (0.631) | (0.049) | (0.048) | (0.001) |
| TEXTILE1 | -0.2224 | -0.2051 | -0.2042 | -0.2229 |
|  | (0.036) | (0.022) | (0.022) | (0.001) |


| Regressor | OLS | Ind. Exp. 2TSF | Corr. Exp. 2TSF | Corr. Exp. 2TSF with Copula |
| :---: | :---: | :---: | :---: | :---: |
| METAL1 | -0.1102 | -0.0611 | -0.0612 | -0.032 |
|  | (0.006) | (0.106) | (0.104) | (0.301) |
| ANYFOR | 0.1483 | 0.1221 | 0.1216 | 0.1173 |
|  | (0.001) | (0.001) | (0.001) | (0.000) |
| ANYSTAT | 0.1002 | 0.1234 | 0.1231 | 0.1189 |
|  | (0.082) | (0.018) | (0.018) | (0.003) |
| LEMP | 0.0558 | 0.057 | 0.0569 | 0.0467 |
|  | (0.002) | (0.001) | (0.001) | (0.001) |
| LVADEMPPPP | 0.1759 | 0.1814 | 0.1818 | 0.1250 |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
| LCAPEMPPPP | 0.0294 | 0.0312 | 0.0312 | 0.0038 |
|  | (0.002) | 0.001) | (0.001) | (0.622) |
| Trend | -0.0885 | -0.0859 | -0.0859 | -0.0527 |
|  | (0.000) | (0.000) | (0.000) | (0.001) |
| Dependent Variable: log monthly earnings. Sample size: 1910. Asymptotic p-values based on heteroskedasticity-robust standard errors (HC2 variant), are given in parentheses. |  |  |  |  |

For many of the coefficients, the estimates are close throughout the four models. But for some of them, and ones that are important in terms for interpretation, the differences are visible: the gender dummy "MALE" appears negative and statistically insignificant in the OLS regression, but ends up large and positive in the 2TSF Correlated Exponential model with a Copula. For Potential Experience ("PEXP"), the opposite happens. The premium for being a Technician ("TECH" dummy) is cut in half between the OLS model and the 2TSF CorrExp with Copula model.

Looking at the employer characteristics, the wage premium for being in the Wood Industry sector ("WOOD 1") is increased 5 times, while being in the Machinery industry ("METAL 1") has now an almost four times smaller effect. Also, the effect of capital intensity (capital/labor ratio-"LCAPEMPPPP") becomes insignificant in the final model.

Turning to the estimates related to the error term and the Copula, we have

Table 4b: Estimates of Error term and Copula parameters

| Residual Skewness <br> Residual Excess Kurtosis | OLS -0.1794 0.5012 | $\begin{array}{r} \text { Ind. Exp. 2TSF } \\ -0.3307 \\ 0.7269 \end{array}$ | Corr. Exp. 2TSF | Corr. Exp. 2TSF with Copula <br> -0.7646 <br> 0.4520 |
| :---: | :---: | :---: | :---: | :---: |
| $\sigma_{\varepsilon}$ | 0.6491 | 0.6566 | 0.6570 | 0.9310 |
| $\sigma_{v}$ | - | 0.3454 | 0.3615 | 0.2203 |
|  |  | (0.000) | (0.000) | (0.000) |
| $\sigma_{w}$ | - | 0.3200 | N/A | N/A |
|  |  | (0.000) |  |  |
| $\sigma_{u}$ | - | 0.4577 | N/A | N/A |
|  |  | (0.000) |  |  |
| $S D(w-u)$ | - | 0.5584 | 0.5486 | 0.9046 |
| $a^{\prime}$ | - | - | 3.1467 | 2.4982 |
|  |  |  | (0.000) | (0.000) |
| $b^{\prime}$ | - | - | 2.1673 | 1.1821 |
|  |  |  | (0.000) | (0.000) |
| $m$ | - | - | 0.4464 | 0.4864 |
|  |  |  | (0.019) | (0.000) |
| Corr(Education, $\varepsilon$ ) | - | - | - | 0.1341 |
|  |  |  |  | (0.000) |
| Corr(Potential Exper, $\varepsilon$ ) | - | - | - | 0.6266 |
|  |  |  |  | (0.000) |
| $E(\varepsilon)=E(w)-E(u)$ | - | -0.1377 | -0.0300 | -0.2059 |
| $\operatorname{Var}(w)-\operatorname{Var}(u)$ | - | -0.1070 | -0.0917 | -0.4440 |

Here, the differences between the four models are more pronounced, and it certainly is the case that the Copula model shares no similarities with the other three, indicating that endogeneity has quantitatively important effects on estimates. For example, in the first three models, the standard deviation of the composite error term is almost identical to $\sigma_{\varepsilon} \approx 0.65$ but in the Copula model it is estimated as 0.93 . Both estimated correlations of the endogenous variables with the error term are statistically significant, providing support for the existence of endogeneity. We note that these correlations are the estimated maximum
correlations for the original, non-transformed variables (see chapter 4). We see that the upper bound for the Education variable is low (0.13), indicating a low level of correlation between length of Education and "Ability". On the other hand the corresponding maximum correlation for the Potential Experience variable is almost five times larger (0.63).

Their positive signs are the anticipated ones: we expect Education to correlate positively with "Ability", and Ability to influence positively the wage. As regards Potential Experience, in generic notation let $x$ denote Potential Experience, and let the relation with Actual Experience $x^{*}$ be $x=x^{*}+e, e \geq 0, \operatorname{Cov}(x, e)>0$. Then the regression relation is

$$
y=\ldots+x^{*} \beta+\varepsilon^{*}=(x-e) \beta+\varepsilon^{*} \Rightarrow y=x \beta+\varepsilon, \quad \varepsilon=-\beta e+\varepsilon^{*} .
$$

So $\operatorname{Cov}(x, \varepsilon)=-\beta \operatorname{Cov}(x, e)$. Since the coefficient on Potential Experience is negative in the model with the Copula, we anticipate $\operatorname{Cov}(x, \varepsilon)>0$, which is what we got. The fact that in the regression specification Potential Experience appears also squared does not affect this qualitative result.

For the test of independence of the one-sided error components $w, u$ developed in chapter 4, we obtained in the Copula model the $\chi_{1}^{2}$-statistic $\hat{q}=11.87$ with $p$-value 0.000 . So the necessary condition for independence is rejected. We also obtain that $\operatorname{sign}\{\operatorname{Cov}(w, u)\}<0$ which is consistent with the theoretical model. The same results were obtained from the 2TSF Correlated Exponential model without a Copula.

In light of the above, we keep the 2TSF Correlated Exponential model with a Copula... and we immediately are confronted with and interesting new issue: we have a regressor, Potential Experience, whose coefficient estimate is economically small and statistically "insignificant", but whose correlation with the error term is estimated as large. What are we to make of this?

Looking at the discussion just above, since we obtained $\beta \approx 0$ it must be the case that Potential Experience covaries strongly with the measurement error $e$ related to actual experience, and so not so much with actual experience per se. Then obtaining virtually zero explanatory power regarding the dependent variable should not come as a surprise: the two
results on Potential Experience are consistent with each other. We consider this a clear indication that the introduction of the Copula allows new insights to be gained.

We now proceed to obtain measures of interest.
In Table 3b we show $E(\varepsilon)=E(w)-E(u)=-0.2059$. This appears to say that the (effects of the) bargaining performance of employers, represented by the variable $u$, is on average much higher than the bargaining performance of the workers, represented by the variable $w$. But we have estimated a semi-log specification, so in order to obtain the effect on the level of the wage we have to calculate

$$
\begin{equation*}
E(\exp \{w-u\})=\frac{(1-m) a^{\prime}}{\left(a^{\prime}-1\right)}+\frac{m b^{\prime}}{\left(b^{\prime}+1\right)}=1.1198 \tag{5.26}
\end{equation*}
$$

We obtain the opposite result, namely that the average net effect on the wage of $w$ and $u$ is an increase of $\sim 12 \%$, indicating that the bargaining effect is in favor of the workers. While the different signs of $E(w-u)$ and $E(\exp \{w-u\})$ may look counter-intuitive, it is nevertheless a possible situation, which has to do with the characteristics of the distributions involved. Their density plots are given below:

Figure 2: Density plots of $z=w-u$ and $e^{z}=e^{w-u}$.



In the first plot, the very long left tail in the negative orthant is what leads to $E(w-u)<0$, even though we have $\operatorname{Pr}(w>u)=1-m=0.514$. But when we exponentiate the difference $w-u$ all the negative values are compacted in the $(0,1)$ interval and with low probability mass for the values close to zero, while the distribution now exhibits a long right tail, resulting in $E(\exp \{w-u\})>1$.

This phenomenon is possible only in a stochastic framework. In a deterministic setting, we would necessarily have $w-u<0 \Rightarrow \exp \{w-u\}<1$. But when expected values get in the way we may obtain $E(w-u)<0, E(\exp \{w-u\})>1$. Put it differently, this is a case where Jensen's inequality holds with extreme prejudice.

This result, that the bargaining effect is on average in favor of the workers, is consistent with the relative scarcity of skilled labor in the manufacturing sector as mentioned in Bigsten et al. (2000) that originally used these data.

The effect of the $v$-component is $E(\exp \{v\})=\exp \left\{\sigma_{v}^{2} / 2\right\}=1.0246$. This captures the effects of updating the ex ante expected productivity. The combined effect is

$$
\begin{equation*}
E(\exp \{\varepsilon\})=E(\exp \{w-u\}) \cdot E(\exp \{v\})=1.1474 \tag{5.27}
\end{equation*}
$$

This means that on average, the negotiation process benefits the workers, allowing them to achieve $\sim 15 \%$ more than what their "typical attributes" would entail as this last magnitude is captured by the estimated common-information expected productivity at the beginning of the negotiation.

But this should be qualified by a "Reality Check" measure: in $52 \%$ of observations, the residual had a negative value, meaning that the agreed wage was below the initial symmetric-information expected productivity. Specifically we have

Table 5: Winners and losers from wage negotiation.

| USD PPP (levels) | $\hat{\varepsilon}<0$ |  |
| ---: | ---: | ---: |
| Conditional mean values | $\hat{\varepsilon}>0$ |  |
| $\exp \left\{\hat{E}\left(p \mid I_{f} \cap I_{e}\right)\right\}$ | 134.91 | 151.27 |
| Actual Monthly Earnings | 78.37 | 260.88 |
| $\% \Delta$ | $-41.91 \%$ | $+72.46 \%$ |
| $\operatorname{Corr}\left[\exp \left\{\hat{E}\left(p \mid I_{f} \cap I_{e}\right)\right\}\right.$, Actual Monthly Earnings] | 0.78 | 0.66 |

Table 5 provides two interesting conclusions: first, the actual wage level depends a lot on the non-typical qualifications of the worker or the characteristics of the firm, since actual wages are on average (far) away from the expected productivity that these qualifications and characteristics would project. Put another way, the bargaining process is critical for the final outcome. Second, that, nevertheless, those with higher professional credentials, as reflected in higher $\hat{E}\left(p \mid I_{f} \cap I_{e}\right)$, do better also in the negotiation stage: the correlation $\hat{E}\left(p \mid I_{f} \cap I_{e}\right)$ and the actual wage is strongly positive.

Finally, since we have pooled intertemporal data, it is of interest to see the evolution of averages through time. In the following table, the first column shows the evolution of the symmetric-information expected productivity prior to negotiations, the second is the net effect of bargaining performance while the third is the net negotiation effect (both as gross markups), and the last column is the actual observed wage.

Table 6: Intertemporal Evolution, sample averages.

|  |  | $\exp \left\{\hat{E}\left(p \mid I_{f} \cap I_{e}\right)\right\}$ | $\hat{E}\left(\exp \left\{w_{i}-u_{i}\right\} \mid \varepsilon_{i}\right)$ | $\exp \left\{\hat{\varepsilon}_{i}\right\}$ <br> (Gross mark-up) | Actual <br> monthly <br> earnings <br> (Gross mark-up) <br> (USD PPP) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1992 | $n$ | (USD PPP) | 1741.98 | 1.047 | 1.065 |
| 1993 | 562 | 149.00 | 1.063 | 1.081 | 171.03 |
| 1994 | 874 | 133.60 | 1.130 | 1.168 | 160.60 |

We observe time trends: expected productivity based on typical qualifications and characteristics tended to fall, while the negotiation stage increased in importance, offsetting to a large degree the previous trend, as evidenced by actual monthly earnings.

We close this empirical application, and the whole chapter, by presenting the intertemporal evolution of negotiation outcomes, expressed as the proportion of positive and negative residuals per year:

Table 7: Negative and positive negotiation outcomes per year.

| Year | $\exp \left\{\varepsilon_{i}\right\}<1$ | $\exp \left\{\varepsilon_{i}\right\}>1$ |
| :---: | :---: | :---: |
| 1992 | $47.3 \%$ | $52.7 \%$ |
| 1993 | $55.7 \%$ | $44.3 \%$ |
| 1994 | $53.0 \%$ | $47.0 \%$ |

Positive negotiation outcomes for the workers fell visibly in the second year/wave of the sample, and rebounded somewhat in the third year. But overall a " $50-50$ " balance is roughly observed, pointing to the unpredictability of the direction of the negotiation outcome.

## Chapter 6

# Re-visiting the production frontier: the contribution of management to production, the "wrong skewness" problem, and a two tier-stochastic frontier model to measure them. 


#### Abstract

We revisit the production frontier of a firm and examine the contribution of management to the firm's output. In order to estimate this contribution without the need to obtain data on management as a production factor, we develop a two-tier stochastic frontier model that can also account for the "wrong skewness" problem or "super-efficiency paradox". The approach contributes to the cost-benefit analysis related to the management system of a company. It can also facilitate research related to management pay and be used in studies of the determinants of management performance. In the first empirical application of the chapter, we contrast the results of the model with the approach of Bloom, Van Reenen \& Associates, where a measure of management is obtained and enters the specification as a regressor. In the second empirical application we estimate the model with a sample that exhibits the wrong skewness issue.


## Introduction.

It is a widely circulated maxim in the business world that, in order to manage something you must be able to measure it. ${ }^{1}$ It is therefore a bit ironic that management itself has, for a long period of time, resisted quantification. And phenomena that are hard to quantify tend to draw a lot of intellectual attention.

To quote from Triebs \& Kumbhakar (2013), "business scholars have long maintained that management is an important factor in production. And it is often perceived to be qualitatively different from conventional input factors and attracts special attention". And not only business scholars, we might add, since the business world's ongoing preoccupation with management is well-known. But economists also have early on expressed an interest on the matter. Bloom, Brynjolfsson, Foster, Jarmin et al. (2017) write that "Economists' interest in management goes at

[^43]least as far back as the 1887 paper 'On the sources of business profits' by Francis Walker", but they also endorse a statement from a recent 2011 study stating that "no driver of productivity has seen a higher ratio of speculation to research." ${ }^{2}$

In this chapter, we aim at lowering this ratio by presenting a 2TSF model of production that includes management as a latent variable, a model that achieves several goals at once: it allows us to estimate the contribution of management to the output of a firm. It permits to separate internal and external inefficiency, and estimate the latter. It permits to examine the relations between management and the conventional inputs. It provides a natural interpretation of the "wrong skewness/super-efficiency" paradox that is often found in production data sets. And, being a latent-variable model, it does not require having data on management beforehand.

We devote section I to the presentation of some alternative approaches to modeling management, that either attempt to measure it directly or to estimate its effects on output based on production and input data. This review serves also the purpose of bringing onto the surface the sometimes subtle methodological questions we have to answer in order to arrive at a consistent mathematical model for management in production. In section II we elaborate on these issues in some detail, and we present our choices. In section III we develop our 2TSF model both from a theoretical perspective and from a statistical one. Section IV contains an empirical application where we contrast our 2TSF model with the approach of Bloom, Van Reenen \& Associates that attempt to measure management directly. Section V discusses the "wrong skewness" issue and contains also an empirical application to showcase how the 2TSF approach solves it naturally while also providing meaningful information on the effects of management on output and its relations with the other inputs.

[^44]
## I. Models of management in production.

## I.1. The Lucas (1978) model.

Even though the management-in-production model developed in Lucas (1978) was not conceived as an econometric tool to estimate the effect of management on output, it stands out because it models management in a two-dimensional way. In the words of the author (p.511) "The managerial technology involves two elements: variable skill or talent, and an element of diminishing returns to scale, or to 'span of control'." In practice, the author defined $f(z)$ to be the "conventional" production function, where $z$ is a vector holding the conventional production inputs, and modeled actual output as $q=x \cdot g(f(z))$ with $0 \leq x \leq 1$ reflecting the relative level of "managerial talent" and $g(\cdot)$ being a strictly increasing concave function that represents the "control system" on the production function $f$. Its concavity reflects the diminishing returns to managerial control.

Here we detect the first methodological question related to management modeling: we can usefully map these two "managerial technology elements" to the distinct concepts of "management" and "leadership". To invoke again a piece of business wisdom, "management is doing things right. Leadership is doing the right things". ${ }^{.}$Indeed, management is associated with the everyday monitoring and control system of a business, while leadership has an intertemporal, mid-to-long term horizon related to the foundational principles and the strategic goals of the company. A researcher has first to decide whether she will model both or which one of them.

Lucas (1978) attempted to model both, imposing specific mathematical properties that necessarily imply specific structural assumptions. Leadership is represented as a relative "quality" $[0,1]$ index. This means that it is bounded in a handicapped interval: The best leadership can do is to offer $g[f(z)]$ which is the maximum possible production given inputs, when "managerial talent" is "highest possible" ( $x=1$ ). But one could assume an unbounded index also.

Moreover, here zero "managerial talent" implies zero output even if the "management system" represented by the $g$ function is in place. This does not appear very realistic: with

[^45]inputs and a management system in place, some production and output can be expected at least in the short run, even if the boss is a disaster and there is no leadership.

Moving to the second "managerial element", the function $g$ represents the management control system imposed on the conventional production function $f$. Assuming diminishing returns to control appears to be a realistic assumption. But the consequent (strict) concavity of $g(\cdot)$ implies that eventually, in the $(f, g(f))$ space the graph of $g$ will cross the $45^{\circ}$ line and we will have $g(f)<f$ from then on. This means that there exists a level of inputs-only production $f^{*}$ above which we would be better off without operating through $x g$ on $f$, even if we had $x=1$. Diagrammatically,

Figure 1. Output in the Lucas (1978) model with maximum managerial talent ( $\mathrm{x}=1$ ).


As long as the conventional output $f$ is below $f^{*}$, actual output $q=g(f)$ is greater than $f$ and the management system boosts output (given leadership). Since the $g$-function is assumed concave, there exists a point A where $q^{*}=f^{*}$ (a fixed point), and management becomes neutral. For values of $f$ above $f^{*}$ like $f^{* *}$ and point B , we have, due to the
concavity of $g, q^{* * *}=g\left(f^{* * *}\right)<f^{* * *}$ and the mere existence of management actually lowers output. So the model incorporates structurally the property that, even with leadership at its best, as scale of operations increases eventually we will reach a point where we would be better off if we scrapped the management system of the company. The issue here is that the detrimental effect of management sets in automatically and unavoidably as scale of operations increases. This does not appear to be a real-world phenomenon and it serves as a useful warning and alert for the traps that await anyone who wants to model management in production.

## I.2. Measuring management: the Bloom, Van Reenen \& Associates research program (BVRA).

A method to directly quantify management (and not indirectly estimate its effect on output) was introduced in Bloom \& Van Reenen $(2006,2007)$. The method they used was (quote) "a practice evaluation tool developed by a leading international management consultancy firm". ${ }^{4}$ It was originally based on telephone interviews with middle managers (plant / shopfloor) where 18 open-ended questions were answered and were subsequently graded by the interviewer on a scale of 1 to 5 (from worst to best). These questions concerned management practices in four main areas (operations, monitoring, targets, incentives ${ }^{5}$ ). The raw "management practice score" was obtained by averaging the scores from all answers of each interviewee. To be used in an empirical study, a "z-score" transformation per question was applied most of the times: per question, we subtract from the raw score the sample average score of the question and we divide by its sample standard deviation. We obtain standardized mean-deviation scores for every answer each individual gave, and then we average over them to obtain the final z-score management index value per interviewee.

The authors also detailed the great lengths they went in order to control for and reduce various sources of bias, and also to validate the method regarding its correlation with the financial performance of the firms involved, in order for it to be acceptable as a valid tool that measures the quality of management practices. Indicatively, interviewers did not know

[^46]which company the interviewed manager worked for, and financial performance of the firm was not discussed during the interview and the management evaluation stage.

This method has since been implemented by its creators using various other data sets. Bloom et al. (2012) used it on nearly 10,000 observations from 20 countries. Bloom, Eifert, Mahajan, McKenzie \& Roberts (2013) investigated management practices in the Indian textile industry. Bloom, Sadun \& Van Reenen (2016) used it with a data sample from 34 countries and more than 10,000 observations, while Bloom et al. (2017) used it on US data to investigate differences at the plant-level.

The underlying economic model treated management as another input factor, while Bloom et al. (2016) discuss the possibility of modeling management as either capital, technology, or a "design" choice (more on these distictions in a while). As regards the econometric implementation, in the original 2007 paper and most of those that followed, the authors used a standard logarithmic linear form with Sales being the dependent variable, and numerous controls alongside conventional production inputs. Since $m_{z}$ in its z -score form is distributed around zero, the constant term of the conventional production function absorbs the effect of the average management score, meaning that what we are estimating here is what happens on output-given-average-management when management deviates from its mean value.

As one could expect, an issue with this method is its cost. In Bloom, Lemos, Sadun, Scur \& Van Reenen (2014) the average cost per interview (i.e. per observation) is estimated at 400 USD (including fixed costs). ${ }^{6}$ In other words, a data set of 2500 observations costs one million USD. For the Bloom et al. (2017) paper, the method has been implemented through a mandatory government survey, signaling that the approach is taking roots in the government institutions and data will be more easily available in the future, at least for the USA. ${ }^{7}$

Moreover, in this last paper the scoring method changed, probably in an attempt to make scoring more objective: here the goal was to obtain a measure of "structured management", how structured is the management system of the firm (p. 7). So the responses

[^47]to the survey's questionnaire were graded according to this criterion. Grading ranged in the [0,1] interval, making " 1 " corresponding to the "most structured" and " 0 " to the "least structured". In the econometric specification the relation in levels $Q=\ldots \exp \left\{\beta_{m} m_{z}\right\}$ remained. Combined, how management is assumed to affect production now changes dramatically. Previously, the approached ended up estimating effects of management due to deviation from its sample mean. Here, a zero management score implies zero effect on output, while $\exp \left\{\beta_{m}\right\}$ represents the absolute upper boundary of the effect that management can have on output.

Triebs \& Kumbhakar $(2013,2018)$ exploited the availability of the BVRA management scores. In the 2013 paper, they specified various models to disentangle time-induced and management-induced technical change over time, essentially attempting to reduce the unexplained part of the latter. In the 2018 paper, the goal is different: here the authors want to test whether productivity estimates are good proxies for the unobserved management, and whether management does indeed operate as a technology, i.e. a production shifter, in a neutral and monotonic way. They used data from the original Bloom \& Van Reenen (2007) study. The results are somehow mixed: productivity estimates correlate positively with the management score, but weakly. So it is valid to use them in order to rank firms in terms of their management, but not for quantifying the latter. Moreover, using semi-parametric techniques, they find that management is not a neutral shifter of the production function: the coefficients characterizing the production function appear to change at different levels of management.

## I.3. Estimating management and its effect on output from production and input data.

We turn now to examine the latent-variable modeling approach to management, where its effect on output is indirectly estimated by the use of production and input data.

An early attempt was Yaron (1960), who used Principal Factors (Components) Analysis in order to create a proxy for the management variable, based on available data series that, he argued, reflected/correlated with the management level of a firm. But he did not get any
meaningful results. Page (1980) proxied the management level using the human capital of managers (education and industry experience).

Mefford (1986) argued that his was a "direct measurement approach", but in practice, he measured management as deviations from preset yearly goals (budget) for output, cost and quality per plant. Dawson \& Hubbard (1987), estimated production and cost functions, and proxied management using a financial performance indicator, as "an ex post indicator of the management input", in their words. Both approaches essentially adopt the "management as results" approach, which is methodologically questionable as we discuss in the next section.

Mundlak (1961) pioneered an individual-effects panel data model in order to estimate the bias on the output elasticities when estimating cross-sectional least-squares regressions while ignoring the existence and effects of management. In the process he showed how one could obtain a data series for the management variable (in mean-deviation form), if panel data are available. The author modeled the management variable as another input in a CobbDouglas function with its own elasticity $Q=B_{0} X^{b_{1}} M^{c}$ (we assume a single additional input for simplicity). Using lower-case letters for variables in logarithms, and adding a random disturbance the panel data specification becomes

$$
\begin{equation*}
q_{i t}=b_{0}+b_{1} x_{i t}+c m_{i}+v_{i t}, \quad i=1, \ldots, N, t=1, \ldots T . \tag{6.1}
\end{equation*}
$$

So the management variable $m_{i}$ was mapped to the time-invariant "individual effect" for each firm, conceptually multiplied by the management output elasticity $c$. Mundlak focused on obtaining unbiased estimates for the slope coefficients. To achieve that he respecified the equation to express the management effect as deviations from its sample mean $\bar{m}$ :

$$
\begin{equation*}
q_{i t}=\left(b_{0}+c \bar{m}\right)+b_{1} x_{i t}+c\left(m_{i}-\bar{m}\right)+v_{i t}, \quad i=1, \ldots, N, t=1, \ldots T . \tag{6.2}
\end{equation*}
$$

This allowed him to apply what is now called the Least-Squares Dummy-Variables (LSDV) estimator, and he obtained estimates for the slope coefficients using the Within
estimator. Then, pooling the data he obtained an estimate for $\left(b_{0}+c \bar{m}\right)$. This gives an estimated series for $c\left(m_{i}-\bar{m}\right)$ (as a product).

The assumption of constant returns to scale in inputs including management, $b_{1}+c=1$ , permitted him to estimate the management output elasticity $\hat{c}=1-\hat{b}_{1}$, and then divide $c\left(m_{i}-\bar{m}\right)$ by $\hat{c}$ to obtain a data series on management in mean-deviation form. The approach had appeared earlier in Hoch (1955), who obtained a measure of "entrepreneurial capacity", essentially advancing the view that the contribution of management to production hides in the regression residuals. ${ }^{8}$

As in the initial BVRA approach, here too the constant term of the regression absorbs the "average level of management". And so, while the obtained estimates allow us to fully rank the participating firms, we cannot estimate the contribution of management to production since we cannot estimate separately its average level. Moreover, the mathematical specification highlights additional important methodological matters: The specification $Q=B_{0} X^{b_{1}} M^{c}$ imposes a monotonic relationship between management and output: more management is always better; less management is always worse, and what's more, zero management implies zero output.

Massell (1967) applied the same econometric technique, but with a cross-sectional sample of multi-product firms (thus having again a two-dimensional sample). Alvarez and Arias (2003), using panel data, modeled management as a fixed individual effect for each firm, and estimated it as a handicap from the most efficient firm in the data, applying the method of Schmidt \& Sickles (1984).

Alvarez, Arias and Greene (2005), in the context of a SF production model, treat also management as a latent input in production, but they specify a translog production function where interaction terms between management and the other inputs appear. As the authors note, this is critical because otherwise the specification would collapse to the "fixed effects" one as implemented in Mundlak (1961). The authors specify also the existence of a

[^48]technically optimal value for management, defined as the value of management that leads to maximum output given the other inputs, and in this way establish a relation between technical inefficiency and the distance of management from this optimal level (their eq. 4). They arrive at a random coefficients model to be estimated with simulated maximum likelihood. What is important to note in their methodology is that the inefficiency component of the error term emerges as a function of the deviation of management from its technically optimal level. Looking at their empirical results, of interest is the fact that similarly to Triebs \& Kumbhakar (2018), they too found evidence that management is not neutral to the structure of the conventional production function.

In an important recent breakthrough, Delis \& Tsionas (2018) develop a latent-variable model to indirectly estimate management using standard cost and production data on firms, and robustly show in different ways that their estimated management scores explain $90 \%$ of the BVRA scores. The authors essentially trade-off the financial cost of creating data sets containing management scores with estimation complexity, since their Bayesian econometric methods are rather involved. But estimation complexity is a cost that is paid in expertise, a currency that has its central bank and issuing authority in academia. The authors validate the work of Bloom, Van Reenen \& Associates, but they provide also an estimation method that sidesteps the need to obtain expensive data on management scores beforehand. Moreover, they offer a convincing argument (p.66) that management can be seen as incorporating and representing all other resources that have been perceived and proposed as influencing production (like human capital, intellectual capital, organizational capital, etc). We fully subscribe to this view, and so the single term representing management in our theoretical specification represents, appropriately in our opinion, all these factors.

## II. Methodological choices in modeling management.

While reviewing the literature in the previous section, we identified a series of conceptual and methodological issues on which a researcher must make a decision in order to model management in a transparent and coherent way. These are:

1) Will we distinguish between management and leadership?
2) Will we treat management as quality or quantity? Relative or not?
3) Will we treat management as capital, technology or as a "design" choice?

4) Is there a technically optimal level of management? Can an "increase in management" reduce output? More generally, is the effect of management on output bounded above?
5) Can management reduce output below its "level without management", and even nullify production?

Answering unambiguously these questions will help the researcher subsequently insert management into a mathematical model that has indeed the properties that align with the assumptions made verbally, something that is not always the case, and it is not guaranteed given the complex nature of the phenomenon under study. ${ }^{9}$

We elaborate on these issues in turn, and we present our choices that will lead to the 2TSF model for management in production.

## II.1. Management and leadership.

For their measuring-management research project, Bloom \& Van Reenen (2007) viewed "management practices" as "... more than the attributes of the top managers: they are part of the organizational structure and behavior of the firm, typically evolving slowly over time even as CEOs and CFOs come and go". Theirs was a deliberate decision to exclude the higher echelons of management, and so also the aspects of mid-term strategy and direction (leadership), concentrating instead on the middle management level, the everyday machinery of running a company.

The above quote hints indirectly to another difference between management and leadership that is critical for their econometric identification and estimation. The authors view the management structure as slow-changing, while at the same time one of their main research goals is to explain the large observed cross-sectional differences in management practices. ${ }^{10}$ So management variability (and therefore the potential of identification) exists mainly along the cross-sectional axis. On the other hand, the authors portray the top-

[^49]management as a come-and-go group that perhaps does not stay long enough in a company to be able to really influence its long-term course. Indeed, "leadership" requires consistency along the time axis to materialize, and not many firms enjoy it. Moreover, leadership is even harder to pin down than management. It appears that the identification prospects for the effects of leadership are weak at best.

Therefore, in our cross-sectional setting, talking about management rather than leadership is more appropriate, and this is what we will do. From a practical point of view, the lowest frequency of production data is one calendar year, which is the definition of the "short-term" in business and accounting language. While internal procedures, controls and managerial personnel in a company can and do change in the course of a year, these affect output slowly so the variability of management effect in a company during a one-year period can be treated as negligible.

## II. 2 Management as quantity or as quality? Absolute or relative?

Management as "quantity" can be understood as a quantification of the extent to which a firm attempts to control its production process, through the structure of decision-making and the delegation of it, and through monitoring and reporting activities. Without making any judgment as to "how well" or "how efficiently" these activities perform, we could quantify them by measuring their sheer number and frequency, as well as the number of decision making levels, whether job descriptions exist and/or how specific and detailed they are, the proportion of managerial positions in the total headcount, etc. Eventually, we would come up with a "quantity index", and not necessarily a relative or bounded one.

If one were to attempt such a measurement exercise, intuitively one would expect an inverted-U relationship between the quantity of management and efficiency, a "Laffer curve for management" where the production process suffers from neglect if control is "too little" while "too much" managerial supervision suffocates the business, and so an intermediate optimal value must exist. This would require particular functional specifications for the management component in the production function.

We offer the following rationale for the existence of diminishing and eventually negative returns to management: management exists because humans exist in the production process, and any desire for direct absolute control is restricted by the exogenous constraint
that humans entertain certain rights that machines do not, rights that seriously restrict the degree of direct control one can exercise over them (to state the obvious if morbid example, when humans are not performing well we cannot open them up to see what is wrong with them, like we can do with a machine). As a consequence, after a point, increasing control can only be attempted through indirect, roundabout ways, which create administrative noise that leads to efficiency losses, and ultimately to reduced output.

This is an argument in favor of the existence of a technically optimal level of management given conventional inputs, without introducing any cost considerations in the picture, and goes together with a "quantity" view of management. We are not aware of any such attempts to measure "management quantity". But it would be the closest we could get in obtaining a measure for management independently of its effects on realized production.

The other approach to directly measure management is to construct a management quality index. The methodological danger here should be obvious: we want to obtain data on management, and then examine its correlation/association with the performance of a company, in order to answer the question "Does, and how much, management quality matter?" But, adopting the "management quality as performance results" approach as some of the papers reviewed earlier did is a fatal methodological flaw since it pre-assigns the association structure between management and business performance whose very existence is what we want to determine. For example, "management quality" should not be measured as some weighted index of "Key Performance Indicators" that many businesses use. ${ }^{11}$ In fact this cautionary remark brings also to the surface the issue of confounding management with productivity and firm heterogeneity.

Bloom et al. (2016) appear to treat the terms interchangeably, writing that "different fields have different labels for management" (p. 2 footnote 3) and providing research examples from trade theory and labor economics that build firm heterogeneity as differentiated productivity into the theoretical model. While productivity is certainly affected by management, this is the "management as results" approach, where we essentially accept a priori that "better results mean better management". The problem with this approach is that

[^50]it silently implies that the success of a business is mostly if not wholly an internal matter, something that all accumulated experience from the business world says that it is not true, since external forces beyond the control of the firm play a critical role both for its success and for its demise. ${ }^{12}$ So we remain skeptical as to whether we should use the concept of productivity in order to model, measure, and learn things about management.
"Firm heterogeneity" on the other hand is a catch-all vague term en vogue, that has been rising in prominence as panel data samples become available and the "individual effects" models of panel data econometrics become more and more applicable. It is the micro-level analogue to the Total Factor Productivity of growth theory, and, as with TFP, "quantifying firms' differences" is only the beginning. The real challenge comes after that stage, when we should attempt to decompose this "unexplained heterogeneity" and attribute as much as possible of it to more specific factors and phenomena that affect the firm's performance, internal or external. The currently available techniques for such decompositions require the use of panel data. ${ }^{13}$

A solution to meaningfully measure management in terms of relative quality is to rely on a yardstick of "general best practices", some broad and abstract consensus coming from market participants and researchers, disassociated from specific firms and even from time periods, a "distilled wisdom" about how management should be set to deal with production in order to have increased chances to succeed. At its root, this approach quantifies "quality in the design of the management system".

To provide a concrete example distinguishing management as quantity and as quality consider the case of a traditional production line. "Management as quantity" would measure how often the line is monitored for defective products, and how often the data collected are processed by a supervisor. Note that here we remain agnostic as to whether there exists some optimal frequency of these monitoring actions, we just count and quantify them. On the

[^51]contrary, management as "quality" would be to judge the recorded frequency of these actions against "accepted best practices", and arrive at a relative "quality index" that would reflect the distance of actual practice from "best" practice.

Various methods to measure the quality of management have been devised in the world of business and consulting. Indicatively, one method has managers viewing videos with case studies and then answering multiple choice questions. Their answers are compared to benchmarks based on numerous previous responses worldwide ${ }^{14}$, categorized per level of management and industries. Here, benchmarking is the way to obtain a scale to measure quality, by essentially invoking the "wisdom of large numbers" (which should not be compared, not even figuratively, with the Law of Large Numbers). The initial scoring method of the research program of Bloom, Van Reenen \& Associates falls also in this category (providing a relative quality index) but the scoring method in the 2016 paper starts to look a bit like a relative quantity index, since measuring how "structured" management practices are is a more neutral measure.

In a latent-variable model, we cannot really support an argument either in favor of a quantity approach, or a quality one, since in practice we cannot impose such a choice on the model. The choice we can make is whether we will attempt to estimate a relative or absolute measure for management. In our 2TSF model we opt for an unbounded management index.

## II.3. Management as capital, technology or a design choice?

In Bloom et al. (2016) a more detailed examination of various ways to model management was undertaken by the authors. They considered three different concepts: Management as technology, management as capital, and the "management as design" approach, where management is contingent, something over which firms optimize given their situation. They noted that the existence of some degree of spillovers regarding management practices led them to adopt the "technology" approach and named their model "Management as Technology" (MAT), but commented that there is no real difference to treating it like capital, offering the rationale that $\mathrm{R} \& \mathrm{D}$ expenses (a central proxy for technology advances) are classified as intangible capital. And indeed they formulated a dynamic model with capital as a state variable with adjustment costs.

[^52]In Economics, the word "capital" is used for something that accumulates and depreciates. Related to management, learning curves and maturity of systems and procedures appear to be natural aspects that "accumulate in value" over time, while personnel turnover and internal re-organizations could obviously mapped to "depreciation".

But certainly management can also be seen as technology, even if only because what management does is to dictate how the other inputs must be combined. Management is a technology, a "soft" technology if you like, the "way of doing business", which may also explain why it is so variable across firms. It is a "recipe", possibly changing all the time as cooking recipes do, but a recipe nevertheless. ${ }^{15}$ But isn't "capital" also a "recipe", a very complex combination of inputs in the actual production process? Still, in econometric models we ignore this character of capital, and we use a value-measure for it. ${ }^{16}$

In Organizational economics a known approach to management is "management as design", where, reasonably, it is argued that firms attempt to optimize their management structure contingent on their environment and goals. ${ }^{17}$ This certainly is based on a fundamental premise of all economic theorizing. Management as "design" is analogous to the "contingency" model in Management science. ${ }^{18}$ This approach explicitly challenges the usefulness of benchmarking for best management practices. As Gibbons and Henderson (2013) write "...many competitively significant management practices cannot be reduced to welldefined action rules that can be specified ex ante and verified ex post. Instead, the implementation of these management practices is critically dependent on context."

Bloom et al. (2016) note that if we combine the "design" approach with a quantification of management, it follows that then, the "management level" could have an interior

[^53]optimum, and not a monotonic relationship with output. In practice though, empirical studies in this area are implemented by formulating questions that have binary answers, leading to the use of dummy variables in a regression setting, and not to a "management quantity" variable. Moreover, we have to mention that here the optimum is not the same concept as the optimum in the management quantity approach discussed earlier. Now the optimization is fully economic: when firms optimize, they always optimize with respect to costs also (and no public relations executive can persuade us otherwise). But this is not especially relevant for a production function model, were we do not explicitly consider economic optimization with respect to the factors of production.

But the main issue is whether we can meaningfully distinguish between these three approaches in a cross-sectional production function model with management as a latent variable.

Treating management as "capital" would amount to treat it in principle as another input. Only in a dynamic model can the stock-variable nature of capital be modeled. Treating management as "variable technology" would not really change the mathematical modeling. Treating management as a "design choice" on the other hand, implying that it is somehow optimized, doesn't change things either, since, this optimization, being economic and not purely technical, does not actually enter a production function model, apart from arguing that what we will indirectly measure represents an optimal (or an attempted optimal) point. But the same can be said for management as capital or as technology. So we see that in practice, we can "announce" management as belonging to any of these categories and it will make no actual difference in the mathematical model, as long as we keep it static.

In fact, the next modeling choices, apparently of a lesser level, are in reality much more critical for the mathematical formulation, the results that we will obtain, and their interpretation.

## II.4. Does a technically optimal level of management exist? Is the effect of management bounded? And should we model these?

We saw previously that if we think of management as a quantity index we can expect an optimal level from a technical perspective, given the other inputs, irrespective of cost considerations. But this is not some special property of management. One of the indisputable
laws of the real world is the diminishing returns to one factor of production while keeping the others constant: the marginal product of an input eventually starts to decline and after a point it becomes negative. Do we usually model this in its full extent?

We don't. Our two basic functional forms, the Generalized Cobb-Douglas and the C.E.S production functions have marginal products that are declining but everywhere positive, and tend to zero only asymptotically. Is there a justification for ignoring an indisputable law in our models?

It appears that there is: we argue that actual production processes are not characterized by inputs in such quantities that their marginal products approach zero, let alone being negative. So both for theoretical but also for empirical work, we only really need functional forms that can represent and approximate production for "middle" regions of input values, with marginal products being safely away from zero. ${ }^{19}$

Then the question we have to answer changes: Do we expect to observe management systems that exceed their technically "optimal" level, and have a negative marginal product? Management consultants will probably answer "yes!" with enthusiasm -but we have to distinguish between unnecessary controls on the one hand, and necessary controls organized and executed badly on the other.

In principle, stochastic frontier analysis should be concerned with technically optimal management even if we are far away from it, in order to determine the production frontier. And there exists a related issue: is the effect of management on output bounded (conditionally on the conventional inputs)? It appears self-evident that it is: given the conventional inputs, management can only do so much. So we face here either an argmax point, or an asymptotic boundary point: to measure either one would be a valuable piece of knowledge.

But in practice the prospects of identification and accurate estimation of these points do not look good: to estimate the former we would need enough firms having management systems with a negative marginal product, and we doubt that such data sets exist. Alternatively, we could envision many firms being close to this technical optimum from the left -but this would ignore the fact that firms actually will optimize over management costs

[^54]also. To estimate a possible asymptotic upper bound on the other hand, we would need many data points where firms have visibly different levels of management but comparable output. This would again ignore the cost-optimization that goes on in the real world.

## II.5. Can management reduce output below its unmanaged level?

A related issue is the effects on output from downward movements of management. A critical observation in this respect is that management is not a necessary condition for production, from a technical point of view. Production requires input factors, knowledge and will. But "knowledge and will" do not reside solely in management but are spread to one or the other degree to all sentient participants in the production process. On the other hand, input factors are a necessary condition for production. Management with zero inputs cannot produce anything. Positive inputs without a management system in place will produce something, through some degree of self-organization and coordination, although in a nonoptimal way. In other words, we argue that "the worst management can do" is to leave production unmanaged, but strictly positive.

One could object by arguing instead that management can directly harm the firm if it is incompetent enough. Such a view appears to conflate management with leadership: indeed the actions of the firm's leaders can in the mid- or long-term literarily sink a company and drive it to extinction. But this does not imply that management will nullify the production capabilities of the firm as represented in its production function. Firms are driven to extinction because they lose demand for their products, or lose control over their costs. But productive ability won't be eliminated by bad management, in a technical sense. ${ }^{20}$ In other words, we argue that the presence of a management system itself, however "inefficient" it may be, is better than no management at all.

[^55]
## III. A 2TSF model of management in production.

Our choices regarding the methodological issues discussed in the previous section are summarized as follows: we concentrate on the management aspect of running a company, and not on leadership, both because we believe that in a cross-sectional setting the effect of management is more pronounced, and because we believe that it is management that can be meaningfully quantified.

Regarding the quantity/quality aspect, we essentially remain agnostic: we will pass the management concept $m$ through a function $h(m)$ that will actually affect output, and it is this function that we are dealing with. This function will be an absolute and not relative measure.

Regarding the capital/technology/design choice aspect, we certainly treat management as a decision variable, while in our cross-sectional setting any dynamic effects of accumulation/depreciation do not appear. But it is modeled as a variable, to match the observed differences in management practices.

Following the same practice as with the conventional inputs, we do not model the existence of a technically optimal level of management, or an upper bound of its effect on output, relying on the presumption that real-world firms are nowhere near such extreme points. We do assume a positive marginal product for management, and that "zero management" still leaves a strictly positive level of output.

## III.1. The production function and the theoretical model.

We define our deterministic theoretical production function as

$$
\begin{equation*}
Q=A F(\mathbf{x}) e^{h(m)}, \quad m \geq 0, h(0)=0, \quad h^{\prime}(m)>0, \quad h^{\prime \prime}(m)<0 . \tag{6.3}
\end{equation*}
$$

Here $A$ is the usual technology constant and $F(\mathbf{x})$ is the conventional production function. We call this part of output as "deterministic" in the sense that it contains the decision variables of the firm (although not only, as we will discuss in a while). Management is defined as a positive number. If it is zero, there exists a strictly positive quantity that can
be produced, the "unmanaged output". As management increases, its effect on output increases due to $h^{\prime}(m)>0$. But the marginal product of management is not necessarily everywhere declining. To have $\partial^{2} Q / \partial m^{2}<0$ we would require not just $h^{\prime \prime}(m)<0$ but the stronger $h^{\prime \prime}(m)+\left(h^{\prime}(m)\right)^{2}<0$.

We want this specification to not just be a tool for econometric estimation, but consistent also with the theory of the firm. Additionally, we want to see what kind of predictions does a theoretical model with such a production function make under optimizing behavior. We therefore consider the cost-minimization problem of a price-taking firm,

$$
\begin{equation*}
\min _{m, \mathbf{x}} C=r_{m} m+\sum_{i=1}^{k} r_{i} x_{i} \quad \text { s.t. } e^{h(m)} F(\mathbf{x})=\bar{Q}, \quad \mathbf{x}=\left(x_{1}, \ldots, x_{k}\right), \tag{6.4}
\end{equation*}
$$

where $\mathbf{x}=\left(x_{1}, \ldots, x_{k}\right)$ is the vector of conventional inputs and where we have subsumed hard technology $A$ in the function $F$. We will omit the functions' arguments for compactness. $F_{i}$ and $F_{i i}$ will denote first partial and second partial derivatives respectively. We make standard assumptions for the conventional inputs, $F_{i}>0, F_{i i}<0, \forall i$.

The first order conditions for this problem are, with $\xi$ being the Lagrange multiplier on the constraint,

$$
\begin{equation*}
r_{m}=\xi h^{\prime} e^{h} F, \quad r_{i}=\xi e^{h} F_{i}, \quad \forall i \tag{6.5}
\end{equation*}
$$

The above implies that at the optimum we will have $h^{\prime}(m)>0$. So even if we had specified $h(\cdot)$ in such a way so as to have an interior maximum, optimizing behavior is expected to keep as away from such points. At the optimum, management quantity will not be at its maximum output effect level, even if such one exists, but at a lower level. This goes some way in rationalizing why complaints about the "unsatisfactory level of control" over a business are heard often and come even from executives themselves -they are perhaps right but they do not take into consideration the cost aspect of the matter. Simply put, maximizing
the effect of management on output is not the optimal thing to do when costs enter the picture. ${ }^{21}$ This supports our methodological choice of ignoring in our modeling the possibility of a technically optimal level of management. ${ }^{22}$

Regarding the second-order sufficient conditions for a minimum we work in the Technical Appendix of the chapter the case with two conventional inputs, and we find that sufficient conditions for the problem to have a well-defined unique minimum is that the $F(\mathbf{x})$ function is concave (for homogeneous functions, this accommodates decreasing and constant returns to scale), and that $h^{\prime \prime}(m) \leq 0$. We stress that we do not need the output function to be concave also in management, i.e. we do not need $\partial^{2} Q / \partial m^{2}=h^{\prime \prime}(m)+\left(h^{\prime}(m)\right)^{2} \leq 0$ for a global minimum in the cost-minimization problem. It suffices that $h^{\prime \prime}(m) \leq 0 .{ }^{23}$

## Complementarity effects at the optimum.

Solving the first-order conditions for the common factors and equating, we obtain the optimal relations

$$
\begin{equation*}
\frac{h^{\prime} F}{r_{m}}=\frac{F_{i}}{r_{i}} \forall i \tag{6.6}
\end{equation*}
$$

From these, we have the implicit relation $G=h^{\prime} r_{i} F-r_{m} F_{i}=0$ and by the implicit function theorem we get

[^56]\[

$$
\begin{equation*}
\frac{d x_{i}}{d m}=-\frac{\partial G / \partial m}{\partial G / \partial x_{i}}=\frac{-h^{\prime \prime} r_{i} F}{r_{K} h^{\prime} F_{i}-r_{m} F_{i i}} . \tag{6.7}
\end{equation*}
$$

\]

Note that we do not assume any causal direction from management to the inputs: we could equivalently examine the inverse relations $d m / d x_{i}$. What we do examine is the tendencies as regards co-movement.

Since $F_{i i}<0$, and given that $h^{\prime}(m)>0$ at the optimum, the denominator is positive. So the sign of the expression depends on the second derivative of the function $h(m)$. If $h^{\prime \prime}(m)=0$ at the optimum we get $d x_{i} / d m=0$. This means that changes in the management level at the optimum will not be accompanied with changes in the conventional inputs. This is not totally unthinkable: we can imagine that, at least in the Services sector, an increase in output may be accommodated by an increase in the management factor only, that will increase the efficiency in the coordination of the other factors without altering their quantity. But it sounds more of a special case rather than the norm. So we come to question the validity of assuming $h^{\prime \prime}(m)=0$ as a structural property. But, for example, the specification in Bloom et al. (2017) $Q=\ldots \exp \left\{\beta_{m} m_{z}\right\}$ has $h\left(m_{z}\right)=\beta_{m} m_{z} \Rightarrow h^{\prime \prime}\left(m_{z}\right)=0$. With $h^{\prime \prime}(m)<0$ the management factor grows with the conventional inputs along the expansion path, which is what we observe in businesses. ${ }^{24}$

Moreover, viewing all inputs together, there are two kinds of effects: one that affects the composition of inputs (allocative effect), and one that affects their joint level leaving their mix unaffected (scale effect). We show in the Technical Appendix that if we assume a twoinput generalized Cobb-Douglas production function, changes in the management level at the optimum should have scale effects only, leaving the input mix unaffected, say:

$$
\begin{equation*}
Q=A K^{a} L^{b} e^{h(m)}, \quad a, b<1 \Rightarrow \frac{d K / d m}{d L / d m}=\frac{K}{L} . \tag{6.8}
\end{equation*}
$$

[^57]Linear expansion paths (for a given price ratio) in a cost-minimization problem are a known property of homothetic production functions. What we showed here is that the inclusion of the management factor in the specific way in an otherwise homothetic production function does not affect this property.

Finally, we note that to estimate the effects of management on output, we don't need to assume a functional form for the management transformation function $h(\cdot)$, except if we want to eventually obtain a (calculated) estimate for $m$. Given the nature of management, it is questionable whether such a magnitude is meaningful in itself, especially in a latentvariable formulation like ours.

## III.2. Internal and external inefficiency.

We enhance equation [6.3] by adding terms for inefficiency and stochastic disturbances,

$$
\begin{equation*}
Q=A F(\mathbf{x}) e^{h(m)} e^{v} e^{-u} \tag{6.9}
\end{equation*}
$$

The variable $e^{v}$ represents positive or negative unanticipated shocks, while $e^{-u}$ represents inefficiency, with $u$ being a non-negative random variable. But which inefficiency are we talking about, now that we have explicitly modeled the management factor? To answer this we must address the relation between management and efficiency in more detail.

We can summarize the goal of management as "a structured effort to optimize the overall efficiency of a firm" (while we could define leadership as the long-term effort to maximize the survival chances and the value of the firm). In their preface, Fried et al. (2008) p. viii, write "Ultimate responsibility for performance rests with management. We believe that inefficiency arises from the varying abilities of managers and that firms with varying degrees of inefficiency that operate in overlapping markets can coexist for some period of time". And in fact it is not uncommon in papers that implement single-tier SF models to refer to the negative inefficiency component of the error term as "managerial inefficiency". This goes as far back as Farrell (1957) who stated that "technical efficiency indicates the gain that can be achieved by simply
'gingering-up' the management". But what Farrel wrote does not assert that inefficiency comes from within -only that it can be reduced through management efforts.

If ultimate responsibility for performance rests with management, then, to achieve performance one must pursue efficiency. Viewing management as the efficiency champion strengthens the motivation for it to be analyzed from an economic point of view, while it does not come from normative considerations but rather from the methodological observation that this is an overarching goal, consistent with any short-term or long-term financial, operational or other objective a firm may have, be it profit-maximization, acquisition of market share, firm-value maximization, quality excellence, a combination of them, or something else. All are served when efficiency increases, all hurt when there are efficiency losses.

But we should not conflate "responsibility for performance" with "liability for all inefficiency". Interpreting the negative error component in SF production models as "managerial inefficiency" ignores the fact that the inefficiency of a firm is also affected by external factors, regulatory regimes, cultural trends and other societal aspects. While it is the responsibility of management to fight also against them, and not accept them passively as given external constraints, still, fully eliminating their negative effects should not be expected, not even reducing them to negligibility. On the other hand, eliminating inefficiency related to how the factors of production cooperate in order to produce is much more under the control of management, and after all, this is indeed what management is really liable for. So not only conceptually, but also in alignment with the relational structure of the real world, we can distinguish inefficiency per its source into internal and external inefficiency. ${ }^{25}$

In a standard SF model, we measure inefficiency given the level of production inputs. In our model, we measure inefficiency given the production inputs and management. Since management is the carrier of internal (in)efficiency, it follows that the variable $m$ in our

[^58]model represents the net management factor, net of internal inefficiency, and so that $e^{-u}$ captures only external inefficiency. ${ }^{26}$

We view the distinction between internal and external inefficiency as having analogies with the "metafrontier" concept. A clear exposition of the metafrontier can be found in O'Donnell, Rao \& Battese (2008), while Amsler, O'Donnell \& Schmidt (2017) provide an enhanced approach. Under the metafrontier concept, we take into account the fact that the feasible sets from which firms choose inputs and technology may not be the same, differing from industry to industry and from country to country. As O'Donnell et al. (2008) write, "...efficiencies measured relative to the metafrontier can be decomposed into two components: a component that measures the distance from an input-output point to the group frontier (the common measure of technical efficiency); and a component that measures the distance between the group frontier and the metafrontier (representing the restrictive nature of the production environment)." The relation to our 2TSF model is that, with the presence of management, external inefficiency captures the loss of efficiency due to the restrictions imposed by the "production environment". The usefulness of the decomposition is scientifically obvious and relevant for policy.

Could we also measure internal inefficiency in the 2TSF model?
We mentioned previously that "actual management" is expressed net of internal inefficiency. But this means that its actual level will deviate from its decided level. Such deviations are not unheard of related to conventional inputs also, especially labor: clocking a specific amount of workhours does not mean that the worker actually worked that much the "efficiency wage" strand of labor market models is based on this "shirking" effect. ${ }^{27}$ But with management, uncertainty is much bigger, exactly because it is a complex but "soft" technology (recipe). Moreover, appealing to universal notions of entropy, we argue that actual management will always deviate below its optimally decided level (we cannot get better management by chance, or by a conspiracy of the stars). Our model will estimate the output effects of actual management. The distance between the decided upon and the actual level of management would represent internal inefficiency.

[^59]In principle, for actual management $m$ we have

$$
m=m_{d}^{*}-d_{m}, \quad 0 \leq d_{m} \leq m_{d}^{*},
$$

where $m_{d}^{*}$ is the optimally decided upon level of management, its chosen frontier (through cost-considerations or other behavioral frameworks), and $d_{m}$ is the deviation from this level, i.e. internal inefficiency. So we have $h(m)=h\left(m_{d}^{*}-d_{m}\right)$. In order to attempt to measure how $m_{d}^{*}, d_{m}$ affect output separately (and separately from the external inefficiency component), we would have to at least free them from $h(\cdot)$, either by assuming a specific form for this function, or by approximating it in a way that achieves econometric identification. As we already mentioned at the end of the previous subsection, we fear that such a step could result in misleading artefacts.

## III.3. Measures of management and inefficiency.

Among the many metrics and measures one could construct in this framework, we present those that we believe convey the most important pieces of information. We view these measures as the random variables that they actually are. This means that they have a distribution, and therefore, we can choose not just the 1st moment (mean) but other characteristics of the distribution (like the mode, or the median) as their most representative value, depending on the context.

## A. Management contribution to output.

Of primary interest and importance is how much of actual output (in percentage terms) can be attributed to management. This can be contrasted with the costs of the management system of the company, in order to assess their performance in a cost-benefit analysis. Unmanaged output is $A F(\mathbf{x}) e^{v} e^{-u}$ so, at the sample level we have

$$
\begin{equation*}
M c \equiv \frac{Q-A F(\mathbf{x}) e^{v} e^{-u}}{Q}=1-\frac{A F(\mathbf{x}) e^{v} e^{-u}}{A F(\mathbf{x}) e^{v+w-u}}=1-e^{-w} . \tag{6.10}
\end{equation*}
$$

By conditioning this on the composite error term we obtain observation-specific measures, again choosing some representative value like the conditional expected value, or the conditional mode.

## B. External inefficiency.

We calculate external inefficiency as the degree to which it drags down output. If external inefficiency was zero, output would be $Q e^{u}=A F(\mathbf{x}) e^{v+w}$.

Then, the expression

$$
\begin{equation*}
E x \operatorname{In} \equiv \frac{Q e^{u}-Q}{Q e^{u}}=1-e^{-u}, \tag{6.11}
\end{equation*}
$$

calculates in absolute terms the percentage reduction of output below what it could have been given all other factors, and if external inefficiency was zero. Note that ExIn $=1-T E$, i.e. one minus the standard measure of Technical Efficiency in single-tier SF production models.

## C. The management output shift-factor.

We also want to know how much management tends to shift output. Setting $h(m) \equiv w$, at the sample level, this is simply

$$
\begin{equation*}
M s \equiv e^{w} \tag{6.12}
\end{equation*}
$$

This is a "gross markup". If we want to obtain the net markup on unmanaged output we subtract unity.

## D. The Technical Efficiency.

In a single-tier SF model,

$$
\begin{equation*}
T E=e^{-u}, \tag{6.13}
\end{equation*}
$$

measures how technically efficient is a firm, as a percentage of full efficiency. In our setting this is the complement of External Inefficiency.

## E. The Efficiency Duel coefficient.

We have modeled management and the external environment as battling it out with the prize being the efficiency of the firm. The measure

$$
\begin{equation*}
E D \equiv e^{w} e^{-u} \tag{6.14}
\end{equation*}
$$

calculates who wins. If it exceeds unity, the management effect overcomes the external inefficiency effect. Subtracting unity gives us the net effect.

## F. Measuring relations between management and conventional inputs.

Once we have an estimated series for management, it is important to examine whether the theoretical predictions we have discussed earlier hold. Even though we will not be measuring $m$ but only $h(m)$, we rely on the result that since $h(m)$ is increasing in $m$, their covariance will be positive. ${ }^{28}$ Also, that the covariance between $h(m)$ and $E[h(m) \mid \varepsilon]$ is positive too.

Suppose therefore that we calculate association measures between the management series and the conventional input series, like Pearson's linear correlation coefficient or more general metrics like Kendall's tau and Spearman's rho that can detect non-linear associations also. The theoretical model developed earlier predicts a positive association. But this prediction is about the economically optimal level of management $m_{d}^{*}$, not actual management $m=m_{d}^{*}-d_{m}$. But we will be estimating, directly or indirectly,

$$
\operatorname{Cov}\left(m, x_{i}\right)=\operatorname{Cov}\left(m_{d}^{*}-d_{m}, x_{i}\right)=\operatorname{Cov}\left(m_{d}^{*}, x_{i}\right)-\operatorname{Cov}\left(d_{m}, x_{i}\right),
$$

[^60]and theory predicts only that $\operatorname{Cov}\left(m_{d}^{*}, x_{i}\right)>0$. Nevertheless, we can reasonably argue that as the management target rises, the deviation from it rises too in absolute terms, so we should have $\operatorname{Cov}\left(d_{m}, x_{i}\right)>0$ also. Then, obtaining $\operatorname{Cov}\left(m, x_{i}\right)>0$ will tell us that the targeted management level correlates more strongly with the conventional input, than does internal inefficiency,
$$
\operatorname{Cov}\left(m, x_{i}\right)>0 \Rightarrow \operatorname{Cov}\left(m_{d}^{*}, x_{i}\right)>\operatorname{Cov}\left(d_{m}, x_{i}\right) .
$$

But if we obtain $\operatorname{Cov}\left(m, x_{i}\right)<0$ we get the opposite result, namely that conventional inputs correlate more strongly with internal inefficiency than with the management target itself.

We argue that $\operatorname{Cov}\left(m, x_{i}\right)>0$ reflects a better situation than the one represented by $\operatorname{Cov}\left(m, x_{i}\right)<0$ as regards managerial efficiency: for if conventional inputs correlate stronger with the distance from the management target, than with the target itself, this indicates, in a relative sense, that
a) firms fail to adjust adequately their management systems to changes in inputs and
b) as a result, the firms have to compensate with higher amounts of conventional inputs in order to offset the deficiencies in the management system.

With a cross sectional data set, we will be able to obtain only sample-average measures here. We would need a panel data set to be able to calculate such covariances per firm.

## III.4. Dependence, regressor endogeneity, and statistical specification.

Dependence between management and external inefficiency would imply that the two non-negative terms in the composite 2TSF error term will be statistically dependent. Conceptually, there are two ways to think about the issue: The one would take the "management as quality" view. Then, to argue that dependence exists would be to argue that externally imposed inefficiency co-varies either with better management (positive correlation) or with worse management (negative correlation). To make it more illustrative, a grossly bureaucratic regulatory environment could "bring the best out of managers" or attract better managers (say, as a "challenge" or even due to a mix of necessity and
stubbornness), pushing management quality up. But it could also be argued that it drives away the good managers and de-motivates those that remain, bringing management quality down. Bloom \& Van Reenen (2010) report a finding of negative correlation between tighter labor market regulations ("higher external constraints") and the subcategory of "incentives" in their management-practices variable, but no correlation on the other subcategories. While this indicates a degree of statistical dependence (because effectively, the regulatory environment "takes over" some degree of control away from management), at the same time it signals that this degree is low and can be ignored in econometric applications. Moreover, their work in general mainly points at competition and ownership structure as the factors being correlated with management quality. Neither of these can be viewed as forming part of the "external inefficiency" component.

The second way would be to take the "management as quantity" view: here, a more comprehensive regulatory environment for example would certainly lead to increased internal monitoring activities, either because the firm has to comply with mandatory reporting, or because it wants to reduce the possibility of fines and/or other frictions with the regulator. And we can reasonably assume that external inefficiency co-varies positively with the degree of regulation. In addition, we may think that in oligopolisitic markets, concerns about industrial espionage (an aspect of the "production environment") will lead to enhanced security protocols (a component of management). So from this point of view, the existence of statistical dependence seems certain, but we consider the induced actions to be incremental to the core monitoring and control activities of management, nevermind that the former may be the favorite discussion subjects in professional circles due to their thriller nature. Overall, and apart perhaps from cases where we examine markets with certain extreme characteristics, we believe that statistical dependence between the two one-sided error components can be ignored without compromising the quality of the econometric results.

But regarding regressor endogeneity, we should certainly account for the possibility of its existence, even if only because we include management in the error component and we assume that management is a decision variable jointly determined with the conventional inputs. ${ }^{29}$ Bloom et al. (2012) report a clear positive linear relationship between their

[^61]management quality variable and firm size (as measured in terms of headcount). On the other hand Bloom et al. (2014) point out that such a correlation gets weaker if external inefficiency constraints are stronger (a reasonable finding).

A final conceptual issue to be tackled is the following: it would seem unrealistic to assume that the management component $h(m)$ follows a distribution that has its mode at zero. This would imply that the most likely situation is that firms do not have management systems in place.

Moreover, once we identify the component $u$ with external inefficiency, it is difficult for it also, to defend the assumption that its most likely value will be near zero: and this is because "external inefficiency" is partly caused by regulatory agencies and other institutions whose very purpose of existence is to impose a compromise between conflicting economic interests, social values and even political ideologies. As such, when these entities create economic inefficiency as we understand it in efficiency analysis, they essentially provide evidence that they are actually fulfilling their goal, that they are doing their job (although the degree of inefficiency compared to the achieved socio-economic compromise should of course be subject to cost-benefit analysis).

We want then a 2TSF specification in which both one-sided components follow distributions that have their mode away from zero -and we got one: the Generalized Exponential, developed in the last section of chapter 3.

In logarithmic form equation [6.9] acquires the structure of the two-tier stochastic frontier model, with the management component $h(m)$ being part of the composite error term:

$$
\begin{equation*}
\ln Q=\ln A+\ln F(\mathbf{x})+\varepsilon, \quad \varepsilon=v+w-u, \quad w \equiv h(m) \tag{6.15}
\end{equation*}
$$

The statistical assumptions are $v \sim \mathrm{~N}\left(0, \sigma_{v}^{2}\right), w \sim \mathrm{GE}\left(2, \theta_{w}, 0\right), u \sim \mathrm{GE}\left(2, \theta_{u}, 0\right)$ and jointly independent, but that the regressors $\mathbf{x}$ are correlated with $\varepsilon$. We will deal with endogeneity by using the Copula modeling and the Gaussian copula density as detailed in chapter 4.

As promised, in order to estimate the model we do not need to have data on management but only on output and the usual production inputs. Also, the 2TSF framework has an additional benefit: it proposes a new interpretation and a way to deal with the "wrong skewness" problem encountered in many samples of production data, an issue that we will take up last. We just note here that the skewness of $\varepsilon=v+w-u$ depends on the difference $\theta_{w}-\theta_{u}$. So the model accommodates the case of positive ("wrong") skewness of the residuals.

Combining the distributional specifications (and their properties, as discussed in chapter 3) with the measures presented earlier, we obtain some interesting statistical images:

First, the gross management output shifter $M s$ and the management contribution to output Mc follow the distribution of the maximum of two i.i.d. random variables (Pareto and Kumaraswamy respectively). So, if we imagine the existence of two notional possible outcomes (say, in a quantum state waiting to be materialized through observation), then we assume that it is the higher of the two that is observed and materializes. Given the relentless push for efficiency and improved financial performance that characterize the business environment, we are inclined to say that this is a realistic aspect of the model.

Second, the usual Technical Efficiency measure (which in our model is the complement to the External Inefficiency), follows the distribution of the minimum of two Beta random variables. Here, if we imagine again the existence of two notional possible outcomes, we assume that we observe the lowest of the two (i.e. the stronger effect of the External Inefficiency). Therefore, the model incorporates the "worst-case scenario" as regards the effects of external inefficiency on the firm's output. Since the workings in the external environment are almost never focused on the material well-being of a firm (except perhaps in cases of "factory towns"), we believe that this also is a prudent and realistic property of the structure we are imposing here.

## IV. Empirical application \#1: the management contribution to output against the "best practices" score of Bloom, Van Reenen \&

 Associates.In the first empirical application of the chapter we contrast the results obtained using the 2TSF model of management in production with the BVRA management scores.

We use a subset of the data used in Bloom et al. (2012), specifically the 1888 observations from the year 2006. ${ }^{30}$ The data are from various countries and industries. In their regressions, the authors use numerous controls, but we will apply the minimum specification. One could question the validity of analyzing such a diverse sample. But comparing average measures over different socioeconomic regimes, regulatory environments and industries, indicates whether any deep structural link exists or not. The regression specification is

$$
\ln S=\beta_{0}+\beta_{1} \ln K+\beta_{2} \ln L+\varepsilon, \quad \varepsilon=v+w-u,
$$

where the dependent variable is Sales. The model is estimated by maximum likelihood. Table 1 contains the estimation results, including the benchmark OLS regression.

[^62]Table 1. Estimation of the Generalized Exponential 2TSF model on production data.

| Sample size: 1888, cross-section. Year: 2006, various countries. Dependent variable: $\ln S$ (logarithm of Sales) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Variable | OLS | Gen Exp 2TSF with Copula | Parameter | Gen Exp 2TSF with Copula |
| Constant | $\begin{array}{r} 4.087 \\ (0.107) \end{array}$ | $\begin{array}{r} 5.672 \\ (0.231) \end{array}$ | $\sigma_{v}$ | $2.07846 \mathrm{e}-06$ |
|  |  |  |  |  |
| lppent | $\begin{gathered} 0.374 \\ (0.022) \end{gathered}$ | 0.104 | $\theta_{w}$ | 0.442 |
|  |  | (0.030) |  | (0.024) |
| lemp | $\begin{array}{r} 0.617 \\ (0.031) \\ \hline \end{array}$ | $\begin{array}{r} 0.816 \\ (0.039) \\ \hline \end{array}$ | $\theta_{u}$ | 0.567 |
|  |  |  |  | (0.021) |
| $\sigma_{\varepsilon}$ | 0.765 | $\begin{aligned} & 0.803 \\ & \text { (calc.) } \end{aligned}$ | $\breve{\rho}($ lppent,$\varepsilon)$ | $\begin{array}{r} 0.355 \\ (0.046) \end{array}$ |
| Residuals skewness <br> Residuals - ex. <br> kurtosis | -0.320 | -0.595 | $\breve{\rho}($ lemp,$\varepsilon)$ | 0.156 |
|  | 5.076 | 4.982 |  | (0.045) |
| Numbers are truncated at 3d decimal digit. Standard errors in parentheses (robustHC2). Description of variables: $\operatorname{lnS}=$ logarithm of Sales, lppent $=$ logarithm of Capital, lemp $=$ logarithm of Labor (number of workers). |  |  |  |  |

The correlation coefficients are positive and accurately estimated. Under multivariate Normality, these are also the maximum, in absolute terms, linear correlation coefficients between the regressors and the error term. As regards the Gaussian specification for the Copula, we have shown in chapter 4 that it is the correct one if the transformed variables included in it follow jointly a multivariate Normal distribution (MVN). We performed the same battery of MVN tests as in chapter 5, using the same online tool. ${ }^{31}$ The results are conflicting: Mardia's skewness-kurtosis test and Henze \& Zinkler's test strongly reject MVN (p-value smaller than 0.01 ) but Royston's test strongly supports MVN (p-value 1.00). In 1888 observations, 62 can be characterized as outliers. As a visual aid, the chi-square $\mathrm{Q}-\mathrm{Q}$ plot is:

[^63]Figure 2: MVN Chi-square Q-Q plot for Capital and Labor.


Note that the axes have different scales. Deviations from multivariate Normality are not many, and even fewer are pronounced. Overall, the results allow us to accept the use of the Copula model and the Gaussian Copula specification.

Turning to the regressor coefficients, contrary to what is usually observed in SF studies, they change visibly between the OLS model and the 2TSF one. ${ }^{32}$

An interesting result is that the variance of the symmetric stochastic disturbance is estimated as essentially zero. This is not a sign of general misspecification. Although the inclusion of the symmetric error disturbance in the single-tier SF model was what initiated stochastic frontier analysis, here we have two other stochastic terms, the one-sided components, and it may very well be the case that in their presence "nothing" is left for a general disturbance term to contribute to the outcome. Moreover the random variable $z=w-u$ has unbounded support, so the associated regularity condition for the properties of

[^64]the maximum likelihood estimator is not violated. ${ }^{33}$ We also note that the residual skewness is mildly negative, which is in accord with the estimates of the theta parameters of the two one-sided components.

Turning to our main interest, in the previous section we have constructed Mc, that measures what percentage of output can be attributed to management. On the other hand, the BVRA management scores measure some concept of "best practices". It is evident that none of our two measures is directly comparable to the BVRA scores. But it would be reasonable to ask the following question: is "better management" reflected in a higher percentage of output being attributed to management? Does higher quality contribute to a larger share of the quantity?

To examine this question, we stratify the sample in quartiles with respect to the BVRA measure and the management contribution measure $E(M c \mid \varepsilon)=1-E\left(e^{-w} \mid \varepsilon\right)$. We then examine the related contingency tables.

For our case where $\varepsilon=w-u$, we can derive the conditional expected values without undue costs. We provide in the Technical Appendix of this chapter the relevant calculations for this special case. We have

$$
\begin{align*}
& E(\exp \{-w\} \mid \varepsilon)= \\
& =\frac{4 \exp \left\{\varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)}\left[\frac{\exp \left\{-\frac{\theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{\theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}-\frac{\exp \left\{-\frac{\theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{\theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}\right] \tag{6.16}
\end{align*}
$$

$$
-\frac{4 \exp \left\{2 \varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)}\left[\frac{\exp \left\{-\frac{2 \theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{2 \theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}-\frac{\exp \left\{-\frac{2 \theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{2 \theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}\right]
$$

[^65]Note that the formula changes depending on whether $\varepsilon$ is greater or lower than zero, as dictated. Statistics and the empirical frequency of this measure are ${ }^{34}$

Figure 3. Empirical relative frequency of $1-E(\exp \{-w\} \mid \varepsilon)$.


We have the following decomposition of the sample:

Table 2. BVRA $-E(M c \mid \varepsilon)$ count table - number of firms.

|  | $E(M c \mid \varepsilon)=1-E\left(e^{-w} \mid \varepsilon\right)$ |  |  |  |  |
| :---: | ---: | ---: | ---: | ---: | ---: |
| BVRA score | $\mathbf{0 . 0 0 - 0 . 2 5}$ | $\mathbf{0 . 2 5 - 0 . 5 0}$ | $\mathbf{0 . 5 0 - 0 . 7 5}$ | $\mathbf{0 . 7 5 - 1 . 0 0}$ | Total |
| $1-2$ | - | 132 | 10 | 2 | $\mathbf{1 4 4}$ |
| $2-3$ | - | 595 | 131 | 26 | $\mathbf{7 5 2}$ |
| $3-4$ | - | 610 | 217 | 51 | $\mathbf{8 7 8}$ |
| $4-5$ | - | 69 | 29 | 16 | $\mathbf{1 1 4}$ |
| Total | $\mathbf{0}$ | 1406 | 387 | $\mathbf{9 5}$ | $\mathbf{1 8 8 8}$ |

[^66]Table 3. Conditional relative frequencies, $P[E(M c \mid \varepsilon) \mid \mathrm{BVRA}]$.

|  | $E(M c \mid \varepsilon)=1-E\left(e^{-w} \mid \varepsilon\right)$ |  |  |  |  |
| :---: | ---: | ---: | ---: | ---: | :---: |
| BVRA score | $\mathbf{0 . 0 0 - 0 . 2 5}$ | $\mathbf{0 . 2 5 - 0 . 5 0}$ | $\mathbf{0 . 5 0 - 0 . 7 5}$ | $\mathbf{0 . 7 5 - 1 . 0 0}$ | Total |
| $1-2$ | - | $91.7 \%$ | $6.9 \%$ | $1.4 \%$ | $100.0 \%$ |
| $2-3$ | - | $79.1 \%$ | $17.4 \%$ | $3.5 \%$ | $100.0 \%$ |
| $3-4$ | - | $69.5 \%$ | $24.7 \%$ | $5.8 \%$ | $100.0 \%$ |
| $4-5$ | - | $60.5 \%$ | $25.4 \%$ | $14.0 \%$ | $100.0 \%$ |
| Marginal rel. frequency | $\mathbf{0}$ | $\mathbf{7 4 . 5 \%}$ | $\mathbf{2 0 . 5 \%}$ | $\mathbf{5 . 0} \%$ | $\mathbf{1 0 0 . 0 \%}$ |

Looking from left to right at each BVRA stratum (row), we observe that in all cases there is a downward tendency, with high relative frequencies in the lowest inhabited quartile. This means that the contribution of management to production has a roughly similar distribution irrespective of the value of the BVRA score.

If the two were strongly positively connected, the left-to-right trend should have been eventually reversed, and in the " $4-5$ " row of the table, where the companies with the best management practices live, the management contribution should have been concentrated to the higher quartiles. But they are not totally unrelated: inspecting top-to-bottom, the relative frequency for the two higher quartiles increases as the BVRA score increases, while the relative frequency for the lowest quartile decreases: there is some reallocation of relative frequency of management contribution from lower to higher quartiles as the BVRA score increases (the distributions become less steep). Also note the interesting fact that for the 4-5 quartile the relative frequency mass "jumps" directly to the highest management contribution quartile.

What do these results tell us? That increasing management quality does not appear to tangibly affect its direct contribution to output, in the majority of cases. But maybe the increases in management quality show in overall efficiency? To examine this we can perform an analogous stratification exercise but now using the efficiency duel measure $E D$ that gives us the net shift-effect of management and external inefficiency on the output level. Since here we have $v \approx 0$ it follows that

$$
E D=\exp \{w-u\}=\exp \{\varepsilon\} \Rightarrow E(E D \mid \varepsilon)=E(\exp \{\varepsilon\} \mid \varepsilon)=\exp \{\varepsilon\} .
$$

This variable does not lie in $[0,1]$. We observe that $\sim 90 \%$ of the values lie below and up to the value " 2 ". So we stratified here in 5 strata per the $E D$ metric. The results are:

Table 4. BVRA $-E(E D \mid \varepsilon)$ count table - number of firms.

|  | $E(E D \mid \varepsilon)=\exp \{\varepsilon\}$ |  |  |  |  |  |
| :---: | ---: | ---: | :---: | :---: | :---: | :---: |
| BVRA score | 0.0-0.5 | $\mathbf{0 . 5 - 1 . 0}$ | $\mathbf{1 . 0 - 1 . 5}$ | $\mathbf{1 . 5 - 2 . 0}$ | $>\mathbf{2 . 0}$ | Total |
| $1-2$ | 70 | 50 | 17 | 5 | 2 | 144 |
| $2-3$ | 201 | 297 | 142 | 47 | 65 | 752 |
| $3-4$ | 110 | 325 | 239 | 99 | 105 | 878 |
| $4-5$ | 16 | 37 | 25 | 11 | 25 | 114 |
| Total | 397 | 709 | 423 | $\mathbf{1 6 2}$ | $\mathbf{1 9 7}$ | $\mathbf{1 8 8 8}$ |

Table 5. Conditional relative frequencies, $P[E(E D \mid \varepsilon) \mid \mathrm{BVRA}]$.

|  | $E(E D \mid \varepsilon)=\exp \{\varepsilon\}$ |  |  |  |  |  |
| :---: | ---: | ---: | ---: | ---: | ---: | :---: |
| BVRA score | $\mathbf{0 . 0 - 0 . 5}$ | $\mathbf{0 . 5 - 1 . 0}$ | $\mathbf{1 . 0 - 1 . 5}$ | $\mathbf{1 . 5 - 2 . 0}$ | $\mathbf{> 2 . 0}$ | Total |
| $1-2$ | $48.6 \%$ | $34.7 \%$ | $11.8 \%$ | $3.5 \%$ | $1.4 \%$ | $100.0 \%$ |
| $2-3$ | $26.7 \%$ | $39.5 \%$ | $18.9 \%$ | $6.3 \%$ | $8.6 \%$ | $100.0 \%$ |
| $3-4$ | $12.5 \%$ | $37.0 \%$ | $27.2 \%$ | $11.3 \%$ | $12.0 \%$ | $100.0 \%$ |
| $4-5$ | $14.0 \%$ | $32.5 \%$ | $21.9 \%$ | $9.6 \%$ | $21.9 \%$ | $100.0 \%$ |
| Marginal rel. frequency | $\mathbf{2 1 . 0} \%$ | $\mathbf{3 7 . 6} \%$ | $\mathbf{2 2 . 4} \%$ | $\mathbf{8 . 6} \%$ | $\mathbf{1 0 . 4 \%}$ | $\mathbf{1 0 0 . 0 \%}$ |

Even with this naive stratification scheme, it is obvious that something important happens when the BVRA score increases: first, note that in the lowest BVRA score band (1-2) the distribution is monotonically declining, but that in the other three the highest relative frequency is in the second $E D$ stratum. Also, for the two highest BVRA score bands, we have $E(E D \mid \varepsilon)>1$ for more than $50 \%$ of the firms.

We do observe that for the higher three BVRA score bands the relative frequencies in the highest $E D$ stratum, $(>2)$ are higher than in the preceding one, but this is most probably the effect of the last $E D$ stratum having greater length. In any case, total probability mass for $E(E D \mid \varepsilon)>1.5$ clearly increases as the BVRA score increases, including the highest BVRA score, where we see an impressive $22 \%$ of the whole stratum to have values $E(E D \mid \varepsilon)>2$.

These results, combined with the previous stratification exercise, tell us that, in line with the general perception, management has organization-wide effects rather than a more distinct contribution to output as happens with the conventional inputs. This also creates some hopes that in a cost-benefit analysis we may find that there is a net "gain" for the other factors from the presence of management.

The results also cross-validate the BVRA approach and the 2TSF model of management in production, since all indications are that, from two totally different roads and using two totally different methodologies they measure the same thing: a positive influence on output beyond conventional inputs, whose variability across firms aligns only with the variability of management practices.

## V. The "wrong skewness" problem in stochastic frontier analysis.

A usual procedure to infer the existence of inefficiency prior to implement a single-tier production SF model on a data set is to run an OLS regression and test the residuals for negative skewness (which would indicate that we have left in the error term a positively skewed random variable that is subtracted from the random disturbance). But what if the sample skewness of the OLS residuals is statistically equal to zero, indicating a symmetric error term? Kumbhakar \& Lovell (2000) comment (p. 73), that then "the data do not support a technical inefficiency story". Worse, what if the sample skewness of the residuals is estimated as being positive? As Greene (2008, p. 115) writes, this would cast doubt on the validity of a
frontier specification. ${ }^{35}$ This is not just a theoretical curiosity, rarely encountered in practice: indicatively, in Caves (1992) 1318 production samples from 5 countries were subjected to a single-tier SF production model. In 355 or $27 \%$ of them the residuals had positive skewness, even after deletion of outliers and the inclusion of control variables (see their Table 1.1, p. 8). This is the "wrong skewness problem", initially called "Type I failure" by Olson, Schmidt and Waldman (1980).

Confronted with the problem, scholars have come up with many different ideas to explain and handle it. One approach looks to statistical properties: Carree (2002) showed that the one-sided error component can exhibit negative skewness (and so its negative which is the one entering the composite error term will exhibit positive skewness), reflecting a sample where low inefficiency has low probability while high inefficiency has high probability, and used the Binomial distribution as a flexible example (since the distribution can exhibit both positive and negative skewness).

Smith (2008), in presenting single-tier SF models where there exists dependence between the (single) one-sided component and the symmetric disturbance, found that positive skewness of the true composite error term may be due to the dependence, implying that in such cases skewness of the OLS residuals as a selection/validation criterion for an SF model may simply be inapplicable. This approach was further elaborated upon by Bonanno, De Giovanni \& Domma (2017).

Rho \& Schmidt (2015) present and extend Waldman's (1982) theoretical results related to maximum likelihood estimation of the one-tier SF model, namely that, under a correct normal specification of the symmetric disturbance, the log-likelihood has a local maximum at a point where the inefficiency parameter is estimated as zero if and only if the skewness of the one-tier composite OLS residual $\hat{\varepsilon}=\hat{v}-\hat{u}$ is strictly positive. ${ }^{36}$ They also present simulated evidence to show that as sample size increases the probability of obtaining positive skewness of the residuals goes to zero, assuming of course that the whole

[^67]specification is correct. In other words, the authors have shown that the "wrong skewness" issue can be explained as a finite-sample artefact possibly strengthened by the existence of fully efficient firms in the sample. Horrace and Wright (2016) generalize Waldman (1982) in an (almost) distributional-free framework, obtaining again that positive skewness theoretically implies zero-inefficiency (and not "super-efficiency"), but they question whether we should accept that this can be the real-world explanation of the wrong skewness issue. Qu , Horrace \& Wu (2015) do see it as a sample issue, and study inequality constrained estimators (MLE and COLS). Simar and Wilson (2009) showed by simulation that when the variance of the inefficiency component is equal to the variance of the stochastic disturbance or lower (i.e. when we have low "signal to noise ratio"), then the wrong skewness appears often in small samples (e.g. with a signal-to-noise ratio equal to unity, 228 samples out of 1000 each of size $n=200$ were characterized by positive skewness). They discussed bootstrap and "bagging" algorithms to obtain meaningful and stable estimates in such situations. Finally, Hafner, Manner \& Simar (2016) assume also that the problem is a smallsample artefact and build a generalized distribution for the one-sided error component in order to circumvent the phenomenon.

The possibility of the existence of "fully efficient" firms has been formally modeled by Kumbhakar, Parmeter \& Tsionas (2013), who developed a model that allows for the probability that some, even many of the firms in the sample are fully efficient. Horrace \& Parmeter $(2014,2015)$, in presenting a stochastic frontier model where the symmetric random disturbance follows a Laplace rather than a Normal distribution, note that among other things the Laplace SF model is conditionally uninformative on firms to which correspond positively valued residuals (which would be the ones to lead to the positive skewness on the whole), and so it permits to focus inference only on those firms with negative residuals, the not-fully-efficient ones.

Almanidis and Sickles (2012) and Almanidis, Qian \& Sickles (2014) presented the "bounded inefficiency" model where the inefficiency component is bounded also from above, and they argue that this has economic justification since grossly inefficient firms are driven out of the market due to the forces of competition. Among the various distributional specifications that they examine, the Doubly Truncated Normal specification can exhibit positive and negative skewness, and so they argue that, since bounded inefficiency is
justified by economic theory, "wrong skewness" is a manifestation of bounded inefficiency with statistical properties that lead to the phenomenon.

Our 2TSF model with management as a latent variable in production advances the "omitted variable" argument: we could obtain a positive skew if we have left in the error term a strong enough (and positively skewed) beneficial component of the production function that exists in all firms but varies in measure for each. The model explains the "wrong skewness" issue by looking for this beneficial factor inwards: this "present in all" but "varying in measure for each" component is the management system of a firm. A positive skew in the OLS residuals is an indication that management, not only "hides in the residuals", but its effect exceeds the negative effects of the exogenous factors.

We note that due to the non-linearity of the logarithmic transformation (that is always applied to production data), a negative skew in the OLS residuals may still co-exist with the management having on average a stronger effect than external inefficiency on the level of output. To give the simplest example, in the benchmark Exponential 2TSF specification (see chapter 3), skewness of the OLS residuals in the logarithmic regression will be negative as long as $E(w)=\sigma_{w}<\sigma_{u}=E(u)$. But this can coexist with $\sigma_{u} /\left(1+\sigma_{u}\right)<\sigma_{w}$ that leads to $E\left(e^{w}\right) E\left(e^{-u}\right)>1$, which means that the net effect on the output level will be positive. When the two one-sided components are assumed statistically dependent the situation may get even more complicated, as we discussed in chapter 4.

Such somewhat counterintuitive parallel properties and results are more pronounced in the 2TSF setting compared to the single-tier one, because here we have two comparable components and we do want to compare them, while in the single-tier SF models there is only one non-negative component. What they do is to weaken the usefulness of skewness as an indication of the relative strength of the latent forces operating on output. Still, this does not change the fact that the 2TSF model with management in production provides a simple and natural explanation of the "wrong skewness" problem.

## V.1. Empirical study \#2: inference in a sample with the wrong skewness

## issue.

We showcase the workings of the 2TSF model by examining a data sample where the "wrong skewness" issue appears. It is a sample of 569 Belgian firms from 1996. ${ }^{37}$ For the production function we used again the Cobb-Douglas logarithmic functional form in capital and labor, and we will implement here too the 2TSF Generalized Exponential specification with a copula density attached as in the previous empirical study.

## V.1.1. Basic estimates and measures at the sample level.

Table 6 contains the estimation results, where we have also included the OLS estimates for comparison. We first note that the correlation coefficients of the (transformed) regressors and the (transformed) error term are estimated high enough and both negative. As regards the Gaussian specification for the Copula, the MVN tests here are: Mardia's skewnesskurtosis test rejected multivariate normality on account of both skewness and kurtosis (pvalues below 0.01). Henze \& Zinkler's test did not reject MVN (p-value 0.107 ) and so did Royston's test (p-value 0.99). As a visual aid, the chi-square $\mathrm{Q}-\mathrm{Q}$ plot shows that the specification can be accepted:

Figure 4: MVN Chi-square Q-Q plot for Capital and Labor.


[^68]In 569 observations only ten of them may be characterized as outliers.
Returning to Table 6, the skewness of the OLS residuals is equal to 1.733 so the sample exhibits the "wrong skewness" issue. The excess-kurtosis is also a characteristic of 2TSF densities, validating the implementation of such a model. We see that the parameter $\theta_{w}$ of the positive one-sided error component is close to double the parameter $\theta_{u}$ of the negative error component, which aligns with the existence of positive skewness. And this is all that needs to be said and done about positive skewness when it exists, when using the 2TSF model: we can proceed with any inference we want, since we have solved the skeweness "paradox" by incorporating it in the fabric of the model.

Table 6. Estimation of the Generalized Exponential 2TSF model on production data.

| Sample size: 569, cross-section. Year: 1996. Country: Belgium. Dependent variable: $\ln \mathrm{Q}$. |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Variable | OLS | Gen Exp 2TSF with Copula | Parameter | Gen Exp 2TSF with Copula |
| Constant | -1.712 | -2.645 | $\sigma_{v}$ | $1.47639 \mathrm{e}-07$ |
|  | (0.179) | (0.188) |  |  |
| $\ln K$ | 0.207 | 0.247 | $\theta_{w}$ | 0.365 |
|  | (0.030) | (0.030) |  | (0.019) |
| $\operatorname{lnL}$ | 0.714 | 0.858 | $\theta_{u}$ |  |
|  |  |  |  | (0.018) |
| $\sigma_{\varepsilon}$ | 0.478 | $\begin{aligned} & 0.464 \\ & \text { (calc.) } \end{aligned}$ | $\breve{\rho}(\ln K, \varepsilon)$ | $\begin{gathered} -0.365 \\ (0.078) \end{gathered}$ |
| Residuals skewness | 1.733 | 1.964 | $\breve{\rho}(\ln L, \varepsilon)$ |  |
| Residuals ex. kurtosis | 10.045 | 11.400 |  | (0.076) |
| Numbers are truncated at 3d decimal digit. Standard errors in parentheses (robust-HC2). Description of variables: $\ln \mathrm{Q}=$ Logarithm of output (value-added), $\operatorname{lnK}=$ logarithm of Capital (total fixed assets, end of 1995), $\operatorname{lnL}=$ logarithm of Labor (number of workers). Monetary values in mil. EUR. |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

We must note that by using Value-Added as the dependent variable (because it was the only output measure available), we ignore the fact that management affects the part of the production process that involves materials/third party costs. At best we make an implicit assumption that materials are roughly a constant fraction of output throughout the sample, a "Leontief" assumption that is not often validated in practice. But we decided to proceed, since the focus of this empirical exercise lies elsewhere.

The regression coefficient for labor changes enough from the OLS estimate. We also observe that the variance of the symmetric error component is estimated as zero. ${ }^{38}$

We turn now to the various measures we have designated previously, whose mode, median and expected value we present in Table 7.

Table 7. Sample-level measures.

| Symbol | Formula | Measure | mode | median | exp. value |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $M c$ | $1-\exp \{-w\}$ | \% of actual output attributed <br> to management | 0.292 | 0.361 | 0.381 |
| $E x I n$ | $1-\exp \{-u\}$ | \% of output lost due to <br> external inefficiency | 0.148 | 0.216 | 0.240 |
| $M s$ | $\exp \{w\}$ | Gross shift factor of <br> management on output | 1.222 | 1.566 | 1.927 |
| $T E$ | $\exp \{-u\}$ | Actual output as \% of output <br> without external inefficiency | 0.852 | 0.784 | 0.760 |
| $E D$ | $\exp \{w-u\}$ | Net shift-effect of <br> management and external <br> inefficiency | 1.020 | 1.205 | 1.464 |

The measure for the percentage of output that can be attributed to management, $M c$ ranges from $29 \%$ to $38 \%$. Analogously, the measure for the percentage of output lost due to the external inefficiency ranges from $15 \%$ to $24 \%$. These two are consistent with the interpretation that positive skewness in the residuals indicates that the positive management force outperforms the negative external one.

[^69]More pronounced differences between the three statistics are to be found in the management shift factor $M s$, where the mode indicates that most likely management tends to shift output by $22 \%$, while on average output almost tends to double due to management. The median value is located approximately in the middle.

The net effect tells a more moderate story: its most likely value is close to unity, i.e. the management system just offsets the external inefficiency. The median value indicates a net shift of $20 \%$ while the average value indicates a net shift of $46 \%$.

Can we assess the plausibility of these estimates? Is it reasonable to say that on average the management system of a firm tends to double "unmanaged output"? Does it look too high -or too low? Is it reasonable to say that $29 \%$ to $38 \%$ of output can be attributed to management? This would be something like the "management output share". Does it appear high -or low? We cannot say, and only continuing research on the matter may provide eventually some stylized facts to be used as benchmarks. Further, which one of these measures are the most useful can only be decided per case and depending on the goals of the study and/or the decisions that need to be taken.

## V.1.2. Individual measures.

In presenting the 2TSF Generalized Exponential specification in chapter 3, we noted that the usual conditional expected value measures are too long and cumbersome, and we opted to develop the conditional modes instead, as a measure at the observation level. But in the current empirical study, the symmetric disturbance $v$ is estimated as having zero variance, and so essentially as being equal to its mean which is zero. This has the following implications: First, we now have $\varepsilon \approx w-u=z$. Therefore, the residuals themselves estimate (consistently) the difference of the one-sided components. Then, the Efficiency Duel metric becomes

$$
E D=\exp \{w-u\}=\exp \{\varepsilon\},
$$

and so we use here the exponentiated residuals themselves. Turning to other individual measures, the expression for $E(\exp \{-w\} \mid \varepsilon)$ has been given previously in eq. [6.16], while

$$
\begin{aligned}
& E(\exp \{w\} \mid \varepsilon)= \\
& =\frac{4 \exp \left\{\varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)}\left[\frac{\exp \left\{-\frac{\theta_{w}+\theta_{u}-\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{\theta_{w}+\theta_{u}-\theta_{w} \theta_{u}}-\frac{\exp \left\{-\frac{\theta_{w}+2 \theta_{u}-\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{\theta_{w}+2 \theta_{u}-\theta_{w} \theta_{u}}\right] \\
& -\frac{4 \exp \left\{2 \varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)}\left[\frac{\exp \left\{-\frac{2 \theta_{w}+\theta_{u}-\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{2 \theta_{w}+\theta_{u}-\theta_{w} \theta_{u}}-\frac{\exp \left\{-\frac{2 \theta_{w}+2 \theta_{u}-\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}}{2 \theta_{w}+2 \theta_{u}-\theta_{w} \theta_{u}}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
& E(\exp \{-u\} \mid \varepsilon)= \\
& =\frac{4 \exp \left\{-\varepsilon / \theta_{w}\right\}}{f_{\varepsilon}(\varepsilon)}\left[\frac{\exp \left\{\frac{\theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \min \{\varepsilon, 0\}\right\}}{\theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}-\frac{\exp \left\{\frac{2 \theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \min \{\varepsilon, 0\}\right\}}{2 \theta_{w}+\theta_{u}+\theta_{w} \theta_{u}}\right] \\
& -\frac{4 \exp \left\{-2 \varepsilon / \theta_{w}\right\}}{f_{\varepsilon}(\varepsilon)}\left[\frac{\exp \left\{\frac{\theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \min \{\varepsilon, 0\}\right\}}{\theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}-\frac{\exp \left\{\frac{2 \theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}{\theta_{w} \theta_{u}} \min \{\varepsilon, 0\}\right\}}{2 \theta_{w}+2 \theta_{u}+\theta_{w} \theta_{u}}\right]
\end{aligned}
$$

With these we can calculate conditional expected values per observation and obtain the main statistics of the resulting series:

Table 8. Statistics of firm-specific measures.

| Variable | Marginal <br> mean | Sample <br> mean | Marginal <br> median | Sample <br> median | Sample <br> SD | Min | Max |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| $E(M c \mid \varepsilon)$ | 0.381 | 0.376 | 0.361 | 0.335 | 0.175 | 0.191 | 0.989 |
| $E(E x \operatorname{In} \mid \varepsilon)$ | 0.240 | 0.239 | $\mathbf{0 . 2 1 6}$ | 0.198 | 0.094 | 0.178 | 0.774 |
| $E(M s \mid \varepsilon)$ | 1.927 | 2.194 | $\mathbf{1 . 5 6 6}$ | 1.544 | 4.752 | 1.268 | 96.848 |
| $E(T E \mid \varepsilon)$ | 0.760 | 0.761 | $\mathbf{0 . 7 8 4}$ | 0.802 | 0.094 | 0.226 | 0.822 |
| $E(E D \mid \varepsilon)$ | $\mathbf{1 . 4 6 4}$ | 1.680 | $\mathbf{1 . 2 0 5}$ | 1.206 | 3.836 | 0.279 | 77.845 |

## 244

For the mean and median we repeat the marginal means and medians we obtained by using the moment generating function of the unconditional distributions (see Table 7).

As regards the mean, for the management contribution $E(M c \mid \varepsilon)$, the external inefficiency $E(\operatorname{ExIn} \mid \varepsilon)$ and the technical efficiency $E(T E \mid \varepsilon)$ metrics, the first two columns are virtually identical. Where we observe visible differences is in the management gross output shifter $E(M s \mid \varepsilon)$ and the efficiency duel measure $E(E D \mid \varepsilon)$. We can attribute these differences to just two outliers (observations \#50, and \#251). If we drop these we obtain the sample means $\hat{E}(M s)=1.952$, and $\hat{E}(E D)=1.484$ which are much closer to the theoretical means (1st column of the table). Certainly these two variables are expected to have long right tails -but the specific two observations are clearly outliers. Looking at the original series, obs \#50 has very large output compared to the reported level of capital and labor, while obs \#251 reports only one employee.

On the other hand, all sample medians of the conditional expected value series are very close to the estimated medians of the marginal distributions. Although the median is robust to outliers, still, this is an interesting result because the conditional distributions are not in general expected to have the same properties as the marginal ones.

Combining the minimum values with the sample medians we observe also that some measures are highly concentrated: as regards the management contribution to output $E(M c \mid \varepsilon)$, half of the firms are squeezed in the interval ( $0.191,0.335$ ). Even more concentrated is the External Inefficiency effect $E(\operatorname{ExIn} \mid \varepsilon)$, where for half of the firms it lies in the narrow interval $(0.178,0.198)$. This is an indication that the external inefficiency is mainly of institutional origin, and so not really varying from firm to firm.

Looking deeper, for the management output shifter $E(M s \mid \varepsilon)$ we find that for $77 \%$ of firms its value does not exceed 2 , meaning that for these firms, management at most tends to double unmanaged output. Regarding the Efficiency Duel measure $E(E D \mid \varepsilon)$, we find that for $29 \%$ of firms its value is smaller than unity, meaning that for these firms management is losing the battle. Then, for the next $\sim 20 \%$ of firms the value of this measure ranges in

## 245

[1.0, 1.2], while for the next $17 \%$ it ranges in [1.2, 1.4], the upper boundary covering $2 / 3$ of all firms.

## Sample stratification.

It is also of interest to stratify the sample. Since management most and foremost is about coordinating and supervising people, we chose to stratify with respect to labor. Specifically, we implemented the European Union main criterion for "Small-medium Enterprises-SMEs" (Recommendation 2003/361/EC), which is the number of employees.

In Table 9 we provide the sub-sample means of our various metrics for this stratification.

Table 9. Sub-sample means of individual measures.

| SME category | \# of <br> employees | \# of firms | $E(M c \mid \varepsilon)$ | $E($ ExIn $\mid \varepsilon)$ | $E(M s \mid \varepsilon)$ | $E(T E \mid \varepsilon)$ | $E(E D \mid \varepsilon)$ |
| ---: | :---: | :---: | ---: | ---: | ---: | ---: | ---: |
| micro | $<10$ | 18 | 0.653 | 0.207 | 8.327 | 0.792 | 6.652 |
| small | $10-49$ | 157 | 0.437 | 0.214 | 2.753 | 0.785 | 2.160 |
| medium | $50-249$ | 305 | 0.345 | 0.246 | 1.709 | 0.753 | 1.280 |
| large | $250-$ | 89 | 0.319 | 0.264 | 1.632 | 0.735 | 1.196 |
| whole sample | $1-10,661$ | 569 | 0.376 | 0.239 | 2.194 | 0.761 | 1.680 |

As the number of employees increases we observe a clear downward trend as regards the contribution of management to output $E(M c \mid \varepsilon)$ and the management shifter $E(M s \mid \varepsilon)$. This finding is not informative about the "quality" of management: what it says is that as firms become larger, conventional inputs appear to increase their influence on determining output. It is an indication that management as an output driver has its limitations, and as firms get bigger the role of management becomes more and more supervisory than anything else.

On the other hand we observe a clear upward trend to the external inefficiency measure $E(E x I n \mid \varepsilon)$, which is reflected in the downward trend of its complement-to-unity $E(T E \mid \varepsilon)$. This also has intuition: we should expect that the socioeconomic environment and its representatives would want to regulate more intensely a firm the larger it gets, and that
other market players would fight harder against a larger firm. These forces create more constraints on the decision-making of a firm, and so they should lead to higher inefficiency. It would be interesting to examine whether these trends hold also for different countries.

## V.1.3. Relation of management to the conventional inputs.

From how we modeled management in the production function in a cost minimization context, we obtained three theoretical results that are the standard ones for the usual inputs also: first, that management is a substitute to the other inputs, for a given level of output. Second, that management does not affect the optimal capital/labor ratio. Third, that along the expansion path management is a complement to the other inputs, i.e. it grows with them. But we also noted that because what we measure is actual management which is below the decided upon level of management, with internal inefficiency occupying the gap, empirical correlation measures may tell a more complex story.

We examine the empirical evidence for the last two of these theoretical results. As a measure of management we use $E[h(m) \mid \varepsilon]=E(w \mid \varepsilon)$, relying on the positive monotonicity of the function $h(m)$, and of the positive covariance between $E[h(m) \mid \varepsilon]$ and $h(m)$. The formula for $E(w \mid \varepsilon)$ when $\varepsilon=w-u$ is

$$
\begin{align*}
E(w \mid \varepsilon) & =\frac{\exp \left\{\varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)} \frac{4}{\theta_{w}+\theta_{u}}\left(\max \{\varepsilon, 0\}+\frac{\theta_{w} \theta_{u}}{\theta_{w}+\theta_{u}}\right) \exp \left\{-\frac{\theta_{w}+\theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\} \\
& -\frac{\exp \left\{\varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)} \frac{4}{\theta_{w}+2 \theta_{u}}\left(\max \{\varepsilon, 0\}+\frac{\theta_{w} \theta_{u}}{\theta_{w}+2 \theta_{u}}\right) \exp \left\{-\frac{\theta_{w}+2 \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\} \\
& -\frac{\exp \left\{2 \varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)} \frac{4}{2 \theta_{w}+\theta_{u}}\left(\max \{\varepsilon, 0\}+\frac{\theta_{w} \theta_{u}}{2 \theta_{w}+\theta_{u}}\right) \exp \left\{-\frac{2 \theta_{w}+\theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}  \tag{6.17}\\
& +\frac{\exp \left\{2 \varepsilon / \theta_{u}\right\}}{f_{\varepsilon}(\varepsilon)} \frac{4}{2 \theta_{w}+2 \theta_{u}}\left(\max \{\varepsilon, 0\}+\frac{\theta_{w} \theta_{u}}{2 \theta_{w}+2 \theta_{u}}\right) \exp \left\{-\frac{2 \theta_{w}+2 \theta_{u}}{\theta_{w} \theta_{u}} \max \{\varepsilon, 0\}\right\}
\end{align*}
$$

We calculate three association measures: Pearson's linear correlation coefficient, as well as Kendall's tau and Spearman's rho that detect non-linear associations also. These are computed for the same strata as before and for the whole sample.

Table 10. Relation between management and the capital/labor ratio.

| $E(w \mid \varepsilon), \quad K / L$ | \# of <br> employees | \# of firms | Pearson | Kendall | Spearman |
| ---: | :---: | :---: | ---: | ---: | ---: |
| micro | $<10$ | 18 | 0.498 | -0.137 | -0.17 |
| small | $10-49$ | 157 | -0.059 | -0.077 | -0.112 |
| medium | $50-249$ | 305 | 0.099 | -0.033 | -0.042 |
| large | $250-$ | 89 | 0.148 | 0.014 | 0.008 |
| whole sample | $1-10,661$ | 569 | 0.286 | -0.009 | -0.011 |
| whole sample excl. 1 outlier | $1-10,661$ | 568 | 0.032 | -0.012 | -0.016 |

With the exception of the Pearson correlation coefficient for the micro-business stratum 0.498 , all the obtained association measures are small (and statistically insignificant). And this one large value is affected by the outliers mentioned previously. The effect of these outliers on the Pearson correlation coefficient is also clear at the full-sample level, where if we drop just the largest of them, the correlation value collapses from 0.286 to 0.032 .

So in all, the sample supports the theoretical result that management does not affect the conventional input mix.

Table 11. Relation between management and inputs along the expansion path.

|  |  |  | $E(w \mid \varepsilon), K$ |  |  | $E(w \mid \varepsilon), L$ |  |  |
| ---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | \# of <br> employees | \# of <br> firms | Pearson | Kendall | Spearman |  | Pearson | Kendall |
| Spearman |  |  |  |  |  |  |  |  |
| micro | $<10$ | 18 | 0.480 | -0.137 | -0.184 | -0.365 | -0.183 | -0.261 |
| small | $10-49$ | 157 | -0.105 | -0.137 | -0.202 | -0.282 | -0.147 | -0.222 |
| medium | $50-249$ | 305 | 0.051 | -0.09 | -0.126 | -0.123 | -0.154 | -0.225 |
| large | $250-$ | 89 | -0.051 | -0.075 | -0.117 | -0.117 | -0.159 | -0.236 |
| whole sample | $1-10,661$ | 569 | -0.042 | -0.199 | -0.285 | -0.102 | -0.248 | -0.357 |

Most of the correlation measures are negative, indicating that along the expansion path of output, actual management level covaries negatively with conventional inputs. Given the
previous discussion on the matter, this is an indication that there is a stronger positive connection between internal inefficiency and conventional inputs, than between inputs and the decided-upon/targeted level of management. Since increases in output and inputs is driven largely from exogenous demand factors, this implies that as the firm's size gets bigger, controlling operations becomes more difficult, internal inefficiency increases, and conventional inputs attempt to compensate.

Regarding capital, we observe two positive values and a downward trend in absolute value, as firms increase in size. Regarding labor, we have negative correlation throughout the strata. Pearson's linear correlation coefficient exhibits a downward trend as the firm size increases, but the general association measures remain roughly at the same level.

Finally, we can see these last results together with the estimated Pearson's correlation coefficients we obtained between each input and the error term through the Copula, also negative. Since here the symmetric error disturbance is treated as non-existent, we have

$$
\operatorname{Cov}\left(x_{i}, \varepsilon\right)=\operatorname{Cov}\left(x_{i}, w-u\right)=\operatorname{Cov}\left(x_{i}, w\right)-\operatorname{Cov}\left(x_{i}, u\right), \quad x_{i}=K, L .
$$

Since we obtained $\operatorname{Cov}\left(x_{i}, w\right)<0$, the fact that the correlations through the Copula are also negative is consistent, but it does not provide any additional insight. If the Copula correlations were positive, then we could say that we have detected strong positive correlation between the inputs and external inefficiency. Of course this correlation could also be checked by deriving $E(u \mid \varepsilon)$ and calculating its sample covariance with the inputs.

With this we conclude the second empirical study of the chapter.

## Summary, extensions and applications.

We have developed a two-tier stochastic frontier model in order to estimate the contribution of the management system of a firm to output, and to account for the "wrong skewness" issue observed in data samples. At the same time, this approach separates internal inefficiency from external inefficiency.

The model treats management as a latent variable and does not require the availability of data on management beforehand. It is able to provide firm-specific estimates related to
management, and allows the examination of the relation between management and conventional inputs.

The first empirical application of the paper explored the correlation of the obtained management contribution to output with the management quality index of Bloom, Van Reenen \& Associates. The association was existent but weak. The second empirical application highlighted the model's workings in a "wrong skewness" situation, indicated that management does not affect the input mix along the expansion path, and also, that conventional inputs correlate more strongly with internal inefficiency than with the targetet management level. Certainly, more empirical studies are needed to assess the usefulness of the model, but the first signs are encouraging.

Extending the 2TSF management in production model into a time-series and/or a panel-data framework is an obvious direction for further development, where one would have to account for any variability of management along the time axis. Furthermore, incorporating management in models that treat other concepts of efficiency, like cost, revenue, or profit efficiency, will provide quantified measures of the management contribution in terms of the concept of efficiency that best aligns with the firm's objectives in each case. Finally, these obtained measures can be usefully applied for purposes of evaluating management pay, and also, as the dependent variables in studies that seek to reveal the determinants of management performance, something that has been noted already by Mundlak (1961).

## References

Aigner D, Lovell CK \& Schmidt P (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6(1):21-37.
Almanidis P \& Sickles RC (2012). The skewness problem in stochastic frontier models: Fact or fiction? In Keilegom IV and Wilson P, eds, Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festchrift for Léopold Simar, 201-227. Springer.
Almanidis P, Qian J \& Sickles RC (2014). Stochastic Frontier Models with Bounded Inefficiency. In Sickles RC \& Horrace WC, eds, Econometric Methods and Applications: Festschrift in Honor of Peter Schmidt, 47-81. Springer. New York.
Alvarez A \& Arias C (2003). Diseconomies of size with fixed managerial ability. American Journal of Agricultural Economics, 85(1):134-142.
Alvarez A, Arias C \& Greene W (2005). Accounting for Unobservables in Production Models: Management and Inefficiency. Efficiency series Working Paper 07/2005, Department of Economics, University of Oviedo.
Amsler C, Prokhorov A \& Schmidt P (2015). Endogeneity in stochastic frontier models. Working Paper BAWP-2015-01. University of Sydney.
Amsler C, O' Donnell CJ \& Schmidt P (2017). Stochastic metafrontiers. Econometric Reviews, 36(6-9):1007-1020.

Atkeson A \& Kehoe PJ (2005). Modeling and measuring organization capital. Journal of Political Economy, 113(5):1026-1053.
Azzalini A (1985). A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12(2):171-178.
Azzalini A \& Capitanio A (2014). The skew-normal and related families. Cambridge University Press. Cambridge UK.
Balakrishnan N \& Lai CD (2009). Continuous Bivariate Distributions, 2nd ed. Springer. Dordrecht.
Bennis W \& Nanus B (1985). Leaders: The strategies for taking charge. Harper \& Row. New York.
Bergstrom T \& Bagnoli M (2005). Log-concave probability and its applications. Economic Theory, 26(2):445-469.
Berndt E, Hall B, Hall R \& Hausman J (1974). Estimation and Inference in Nonlinear Structural Models. Annals of Economic and Social Measurement, 3(4):653-665.
Bigsten A, Isaksson A, Söderbom M, Collier P, Zeufack A, Dercon S, Fafchamps M, Gunning JW, Teal F, Appleton S, Gauthier B, Oduro A, Oostendorp R, Pattillo C (2000). Rates of return on physical and human capital in Africa's manufacturing sector. Economic Development and Cultural Change, 48(4):801-827.
Binmore K, Rubinstein A \& Wolinsky A (1986). The Nash bargaining solution in economic modelling. The RAND Journal of Economics, 17(2):176-188.
Birnbaum ZW (1942). An inequality for Mill's ratio. The Annals of Mathematical Statistics, 13(2):245-246.
Blanco G (2017). Who benefits from job placement services? A two-sided analysis. Journal of Productivity Analysis, 47(1):33-47.
Bloom N \& Van Reenen J (2006). Measuring and Explaining Management Practices across Firms and Countries. Discussion Paper 716, Center for Economic Performance.
Bloom N \& Van Reenen J (2007). Measuring and Explaining Management Practices across Firms and Countries. The Quarterly Journal of Economics, 122(4):1351-1408.
Bloom N \& Van Reenen J (2010). Why do management practices differ across firms and countries? The Journal of Economic Perspectives, 24(1):203-224.
Bloom N, Genakos C, Sadun R \& Van Reenen J (2012). Management Practices across Firms and Countries. Working Paper No. w17850. National Bureau of Economic Research.

Bloom N, Eifert B, Mahajan A, McKenzie D \& Roberts J (2013). Does management matter? Evidence from India. The Quarterly Journal of Economics, 128(1):1-51.
Bloom N, Lemos R, Sadun R, Scur D \& Van Reenen J (2014). The New Empirical Economics of Management. Working Paper No. w20102. National Bureau of Economic Research.
Bloom N, Sadun R \& Van Reenen J (2016). Management as a Technology? Working Paper No. w22327. National Bureau of Economic Research.
Bloom N, Brynjolfsson E, Foster L, Jarmin RS, Patnaik M, Saporta-Eksten I \& Van Reenen J (2017). What drives differences in management? Discussion Paper w23300. National Bureau of Economic Research.
Bonanno G, De Giovanni D \& Domma F (2017). The 'wrong skewness' problem: a re-specification of stochastic frontiers. Journal of Productivity Analysis, 47(1):49-64.
Brascamp HJ \& Lieb EH (1975). Some Inequalities for Gaussian Measures and the Long-range Order of the One-Dimensional Plasma. In A.M. Arthurs, ed., Functional Integration and Its Applications, Oxford: Clarendon Press. New York.
Cameron AC \& Trivedi PK (2005). Microeconometrics: methods and applications. Cambridge University Press. Cambridge.
Carree MA (2002). Technological inefficiency and the skewness of the error component in stochastic frontier analysis. Economics Letters, 77(1):101-107.
Caves RE (ed) (1992). Industrial Efficiency in Six Nations. MIT Press. Cambridge, Massachusetts.
Chawla M (2002). Estimating the extent of patient ignorance of the health care market. In Devarajan S. \& Halsey Rogers F, eds, World Bank Economists' Forum vol.2, 3-24. World Bank.
Chen YY \& Wang HJ (2004). A method of moments estimator for a stochastic frontier model with errors in variables. Economics Letters, 85(2):221-228.
Chiang AC (1984). Fundamental methods of mathematical economics, 3d ed. McGraw Hill. Auckland.
Coelli T (1995). Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. Journal of Productivity Analysis, 6(3):247-268.
Colombi R, Kumbhakar SC, Martini G \& Vittadini G (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. Journal of Productivity Analysis, 42(2):123-136.
Cornwell C \& Rupert P (1988). Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. Journal of Applied Econometrics, 3(2):149-155.
Covey SR (1989). The 7 habits of highly effective people. Free Press. New York.
Danaher PJ \& Smith MS (2011). Modeling multivariate distributions using copulas: Applications in marketing. Marketing Science, 30(1):4-21.
Dar A (2014). The Impact of Imperfect Information on the Wages of Native-Born and Immigrant Workers: Evidence from the 2006 Canadian Census. Proceedings of 2nd Economics E Finance Conference, Vienna, 215-232. International Institute of Social and Economic Sciences.
Das T \& Polachek SW (2017a). Estimating Labor Force Joiners and Leavers Using a Heterogeneity Augmented Two-Tier Stochastic Frontier. IZA Discussion Paper \#10534.
Das T \& Polachek SW (2017b). Estimating Labor Force Joiners and Leavers Using a Heterogeneity Augmented Two-Tier Stochastic Frontier. Journal of Econometrics, 199(2):93-232.
Dawson PJ \& Hubbard LJ (1987). Management and size economies in the England and Wales dairy sector. Journal of Agricultural Economics, 38(1):27-38.
Delis MD \& Tsionas MG (2018). Measuring management practices. International Journal of Production Economics, 199:65-77.
Drucker PF (1986). Management: Tasks, Responsibilities, Practices. Truman Tulley Books. New York.
Du Z \&Wei P (2014). Agglomeration economy, external governance and the efficiency of urban industrial emissions. Urban Problems, 2014-10.
El Mehdi R \& Hafner CM (2014). Inference in stochastic frontier analysis with dependent error terms. Mathematics and computers in simulation, 102:104-116.


Egozcue $M$ (2015). Some covariance inequalities for non-monotonic functions with applications to mean-variance indifference curves and bank hedging. Cogent Mathematics 2:991082 http://dx.doi.org/10.1080/23311835.2014.991082.
Erdelyi A, Magnus W, Oberhettinger F \& Tricomi FG, eds, (1954). Tables of Integral Transforms: Vol.1. McGraw-Hill. New York.
Erickson T, Jiang CH \& Whited TM (2014). Minimum distance estimation of the errors-in-variables model using linear cumulant equations. Journal of Econometrics, 183(2):211-221.
Farrell MJ (1957). The Measurement of Production Efficiency. Journal of the Royal Statistics Society A, 120(3):253-290, with discussion.
Ferona A \& Tsionas EG (2012). Measurement of excess bidding in auctions. Economics Letters, 116(3):377-380.
Fisher RA (1930). Moments and product moments of sampling distributions. Proceedings of the London Mathematical Society, 2(1):199-238.
Freedman D \& Diaconis P (1981). On the histogram as a density estimator: L2 theory. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 57(4):453-476.
Freund JE (1961). A bivariate extension of the exponential distribution. Journal of the American Statistical Association, 56(296):971-977.
Fried HO, Lovell CK \& Schmidt SS, eds (2008). The measurement of productive efficiency and productivity growth. Oxford University Press. Oxford.
Gaynor M \& Polachek SW (1994). Measuring information in the market: An application to physician services. Southern Economic Journal, 60(4):815-831.
Genest C, Ghoudi K \& Rivest LP (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82(3):543-552.
Genest C \& Nešlehová J (2007). A primer on copulas for count data. ASTIN Bulletin: The Journal of the IAA, 37(2):475-515.
Genest C, Nešlehová JG \& Rémillard B (2014). On the empirical multilinear copula process for count data. Bernoulli, 20(3):1344-1371.
Genton MG \& Loperfido NM (2005). Generalized skew-elliptical distributions and their quadratic forms. Annals of the Institute of Statistical Mathematics, 57(2):389-401.
Gibbons R \& Henderson R (2013). What Do Managers Do? in Gibbons R \& Roberts J, eds, The Handbook of Organizational Economics. Princeton University Press. Princeton.
Godfrey LG \& Orme CD (1991). Testing for skewness of regression disturbances. Economics Letters, 37(1):31-34.
Golub GH \& Welsch JH (1969). Calculation of Gauss Quadrature Rules, Mathematics of Computation, 23:221-230.
Gordon RD (1941). Values of Mills' ratio of Area to Bounding Ordinate and of the normal probability integral for large values of the argument. The Annals of Mathematical Statistics, 12(3):364-366.
Gradshteyn I \& Ryzhik IM (2007). Tables of integrals, sums, series and products, 7th ed. Academic Press. Amsterdam.
Greene WH (1985). Frontier production functions: Comment. Econometric Reviews, 4(2):335-338.
Greene WH (1990). A gamma-distributed stochastic frontier model. Journal of Econometrics, 46(1-2):141163.

Greene WH (2008). The Econometric Approach to Efficiency Analysis. In: Fried HO, Lovell CK, Schmidt SS (eds) The measurement of productive efficiency and productivity growth. Oxford University Press. Oxford.
Greene W (2010). A stochastic frontier model with correction for sample selection. Journal of Productivity Analysis, 34(1):15-24.
Groot W \& Oosterbeek H (1994). Stochastic reservation and offer wages. Labour Economics, 1(3):383390.

Groot W \& van den Brink HM (2007). Optimism, pessimism and the compensating income variation of cardiovascular disease: A two-tiered quality of life stochastic frontier model. Social Science $\mathcal{E}$ Medicine, 65(7):1479-1489.
Gumbel EJ (1960). Bivariate exponential distributions. Journal of the American Statistical Association, 55(292): 698-707.
Gupta RD \& Kundu D (1999). Theory \& methods: Generalized exponential distributions. Australian $\mathcal{E}$ New Zealand Journal of Statistics, 41(2):173-188.
Hadley L \& Ruggiero J (2006). Final-Offer Arbitration in Major League Baseball: A Nonparametric Analysis. Annals of Operations Research, 145(1):201-209.
Hafner CM, Manner H. \& Simar L (2016). The "wrong skewness" problem in stochastic frontier models: A new approach. Econometric Reviews, 37(4):380-400.
Hajek J, Sidak Z \& Sen PK (1999). Theory of rank tests, 2nd ed. Academic Press. San Diego, USA.
Hall M \& Winsten C (1959). The ambiguous notion of efficiency. The Economic Journal, 69(273):71-86.
Hallin M \& Ley C (2012). Skew-symmetric distributions and Fisher information-a tale of two densities. Bernoulli, 18(3):747-763.
Hausman JA (1978). Specification tests in econometrics. Econometrica, 46(6):1251-1271.
Henze N (2002). Invariant tests for multivariate normality: a critical review. Statistical Papers, 43(4):467506.

Hoch I (1955). Estimation of production function parameters and testing for efficiency. In Report of the Montreal Meeting, September 10-13, 1954. Econometrica, 23(3):325-326.
Hofler R \& SW Polachek (1982). Ignorance in the Labor Market: A New Approach for Measuring Information Content. Proceedings of the American Statistical Association, 422-425.
Horrace WC \& Parmeter CF (2014). A Laplace Stochastic Frontier Model. Working paper No. 166. Center for Policy Research, Maxwell School, Syracuse University.
Horrace WC \& Parmeter CF (2015). A Laplace stochastic frontier model. Econometric Reviews, DOI: 10.1080/07474938.2015.1059715.

Horrace WC \& Wright IA (2016). Stationary Points for Parametric Stochastic Frontier Models. Suracuse University. Center for Policy Research. 228. https://surface.syr.edu/cpr/228.
Huang S, Jia J \& Wang W (2017). Research of IPO Underpricing Decomposition Based on Two-tiered Stochastic Frontier Model: Evidence from Chinese Growth Enterprises Market. CMS, 25(2):2129.

Huang Z (2013). The Study of IPO Pricing Efficiency in Chinese GEM Market : An Empirical Measure based on Two-tier stochastic Frontier Model. Journal of Guangdong University of Finance and Economics (2). http://en.cnki.com.cn/Article_en/CJFDTOTAL-SONG201302006.htm.
Huang Y, Luo S \& Wang H (2018). Asymmetric Business Cycles and Economic Uncertainties: An Application of Two-Tier Stochastic Frontier Model. Manuscript.
Jamalizadeh A, Pourmousa R \& Balakrishnan N (2009). Truncated and limited skew-normal and skew-t distributions: properties and an illustration. Communications in Statistics-Theory and Methods, 38(16-17):2653-2668.
Jondrow J, Knox Lovell CA, Materov IS \& Schmidt P (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19(2):233-238.
Jones MC (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. Statistical Methodology, 6(1):70-81.
Kaldor N (1934). The equilibrium of the firm. The Economic Journal, 44(173):60-76.
Kendall MG (1945). The Advanced Theory of Statistics vol 1, 2nd ed. C. Griffin and Co. London.
Kerstens K, O'Donnell C \& Van de Woestyne I (2015). Frontier metatechnologies and convexity: a restatement. Working Paper 2015-08. CNRS-LEM.
Kinukawa S \& Motohashi K (2010). Bargaining in Technology Markets: An empirical study of biotechnology alliances. RIETI Discussion Paper Series 10-E-200.


Kinukawa S \& Motohashi K (2016). What determines the outcome of licensing deals in market for technology? Empirical analysis of sellers and buyers in biotechnology alliances. International Journal of Technology Management, 70(4):257-280.
Klaassen CA \& Wellner JA (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. Bernoulli, 3(1):55-77.
Koop G \& Tobias JL (2004). Learning about heterogeneity in returns to schooling. Journal of Applied Econometrics, 19(7):827-849.
Kopp RJ \& Mullahy J (1990). Moment-based estimation and testing of stochastic frontier models. Journal of Econometrics, 46(1-2):165-183.
Korkmaz S, Goksuluk D \& Zararsiz G (2014). MVN: an R package for assessing multivariate normality. The $R$ Journal, 6(2):151-162.
Kumaraswamy P (1980). A generalized probability density function for double-bounded random processes. Journal of Hydrology, 46(1-2):79-88.
Kumbhakar SC \& Knox Lovell CA (2000). Stochastic Frontier Analysis. Cambridge University Press. Cambdridge.
Kumbhakar SC \& Parmeter CF (2009). The effects of match uncertainty and bargaining on labor market outcomes: evidence from firm and worker specific estimates. Journal of Productivity Analysis, 31(1):1-14.
Kumbhakar SC \& Parmeter CF (2010). Estimation of hedonic price functions with incomplete information. Empirical Economics, 39(1):1-25.
Kumbhakar SC, Parmeter CF \& Tsionas EG (2013). A zero inefficiency stochastic frontier model. Journal of Econometrics, 172(1):66-76.
Kumbhakar SC, Lien G \& Hardaker JB (2014). Technical efficiency in competing panel data models: a study of Norwegian grain farming. Journal of Productivity Analysis, 41(2):321-337.
Kumbhakar SC, Parmeter C \& Zelenyuk V (2018). Stochastic Frontier Analysis: Foundations and Advances. CEPA Working Papers Series WP02/2018. School of Economics, University of Queensland, Australia.
Lakhdar B, Leleu CH, Vaillant NG, and Wolff FC (2013). Efficiency of Purchasing and Selling Agents in Markets with Quality Uncertainty: The Case of Illicit Drug Transactions, European Journal of Operational Research, 226(3):646-657.
Lehmann EL (1999). Elements of large-sample theory. Springer. New York.
Li C, Wang Y \& Zheng Z (2014). The Path Analysis of the Dual Principal-agent Problem Affects the Over-investment of Listed Companies: Based on the Two-tier Heterogeneity Stochastic Frontier Model CMS, 22(11): 131-139. https://doi.org/10.16381/j.cnki.issn1003-207x.2014.11.014.
Lian Y \& Chung CF (2008). Are Chinese Listed Firms Over-Investing? Manuscript. http://dx.doi.org/10.2139/ssrn. 1296462.
Lin ZJ, Liu S \& Sun F (2017). The impact of financing constraints and agency costs on corporate R\&D investment: Evidence from China. International Review of Finance, 17(1):3-42.
Lins MPE, de Lyra Novaes LF \& Legey LFL (2005). Real Estate Appraisal: A Double Perspective Data Envelopment Analysis Approach, Annals of Operations Research, 138(1):79-96.
Liu H (2017). Dynamic Identification and Early-warning of Inefficient Investment Carried out by Listed Companies Based on the Stochastic and PVAR Model. Boletín Técnico, 55(8):708-717.
Liu X \& Liu F (2014). IPO Pricing Efficiency Measurement on Listed SMEs in China: Under the Condition of Asymmetric Information. Review of Investment studies (06). http://en.cnki.com.cn/Article_en/CJFDTotal-TZYJ201406010.htm
Lombera DC (2007). Analysis of Two-party Salary negotiations: Factors impeding a mutuallybeneficial joint agreement. Master Thesis. Getúlio Vargas Foundation.
Lucas Jr RE (1978). On the size distribution of business firms. The Bell Journal of Economics, 9(2):508-523.
Lv X (2013). Agency Conflicts, Incentive Compatibility and Technology Innovation of Chinese Listed Companies: Empirical Analysis Based on Two-tier Stochastic Frontier Model. Modern Finance and Economics - Journal of Tianjin University of Finance and Economics (11).
http://en.cnki.com.cn/Article_en/CJFDTOTAL-XCXB201311012.htm
Lyu X, Decker C \& Ni J (2018). Compensation negotiation and corporate governance: the evidence from China, Journal of Chinese Economic and Business Studies, DOI:10.1080/14765284.2018.1445081.
Ma Y, Genton MG \& Parzen E (2011). Asymptotic properties of sample quantiles of discrete distributions. Annals of the Institute of Statistical Mathematics, 63(2):227-243.
Mari DD and Kotz S (2001). Correlation and dependence. Imperial College Press. London.
Mas-Colell A, Whinston MD \& Green JR (1995). Microeconomic theory (Vol. 1). Oxford University Press. New York.
Massell BF (1967). Elimination of management bias from production functions fitted to cross-section data: a model and an application to African agriculture. Econometrica, 35(3-4):495-508.
Meeusen W \& van Den Broeck J (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18(2):435-444.
Mefford RN (1986). Introducing management into the production function. The Review of Economics and Statistics, 68(1):96-104.
Mohammadi M (2016). On the bounds for diagonal and off-diagonal elements of the hat matrix in the linear regression model. REVSTAT-Statistical Journal, 14(1):75-87.
Moran PAP (1967). Testing for correlation between non-negative variates. Biometrika, 54(3-4):385-394.
Mouchart M \& Vandresse M (2007). Bargaining Powers and Market Segmentation in Freight Transport. Journal of Applied Econometrics, 22(7):1295-1313.
Mouchart M \& Vandresse M (2010). A double-frontier approach for measuring market imperfection. Annals of Operations Research, 173(1):137-144.
Mundlak Y (1961). Empirical production function free of management bias. Journal of Farm Economics, 43(1):44-56.
Murphy A \& Strobl E (2008). Employer and Employee Ignorance in Developing Countries: The Case of Trinidad and Tobago. Review of Development Economics, 12(2):339-353.
Nash Jr JF (1950). The Bargaining Problem. Econometrica, 18(2):155-162.
Nash Jr JF (1953). Two-person cooperative games. Econometrica, 21(1):128-140.
Nelsen RB (2006). An introduction to copulas, 2nd ed. Springer Science \& Business Media. New York.
O'Donnell CJ, Rao DP \& Battese GE (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empirical economics, 34(2):231-255.
Olley G \& Pakes A (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. Econometrica, 64(6):1263-1297. doi:10.2307/2171831.
Olson JA, Schmidt P \& Waldman DM (1980). A Monte Carlo study of estimators of stochastic frontier production functions. Journal of Econometrics, 13(1):67-82.
Owen DB (1956). Tables for computing bivariate normal probabilities. The Annals of Mathematical Statistics, 27(4):1075-1090.
Owen DB (1980). A table of normal integrals. Communications in Statistics - Simulation and Computation, 9(4):389-419.
Page JM (1980). Technical efficiency and economic performance: Some evidence from Ghana. Oxford Economic Papers, 32(2):319-339.
Papadopoulos A (2015a). The half-normal specification for the two-tier stochastic frontier model. Journal of Productivity Analysis, 43(2):225-230.
Papadopoulos A (2015b). The Double Half-normal distribution and its extensions. Technical Report. Athens University of Economics and Business.
Park S \& Gupta S (2012). Handling endogenous regressors by joint estimation using copulas. Marketing Science, 31(4):567-586.
Parmeter CF (2017). Estimation of the Two-Tiered Stochastic Frontier Model with the Scaling Property. Journal of Productivity Analysis, DOI 10.1007/s11123-017-0520-8.
Parzen E (2004). Quantile probability and statistical data modeling. Statistical Science, 19(4):652-662.
Pissarides CA (2000). Equilibrium unemployment theory, 2nd ed. MIT Press. Cambridge, Massuchusetts.

Poggi A (2010). Job satisfaction, working conditions and aspirations. Journal of Economic Psychology, 31(6):936-949.
Polachek SW \& Yoon BJ (1987). A two-tiered earnings frontier estimation of employer and employee information in the labor market. The Review of Economics and Statistics, 69(2):296-302.
Polachek SW \& Yoon BJ (1996). Panel estimates of a two-tiered earnings frontier. Journal of Applied Econometrics, 11(2):169-178.
Polachek SW \& Robst J (1998). Employee Labor Market Information: Comparing Direct World of Work Measures of Workers' Knowledge to Stochastic Frontier Estimates. Labour Economics, 5:231-42.
Polachek S, Das T \& Thamma-Apiroam R (2013). Heterogeneity in the Production of Human Capital. IZA Discussion Paper \#7335.
Polachek S, Das T \& Thamma-Apiroam R (2015). Micro- and Macroeconomics Implications of Heterogeneity in the Production of Human Capital. Journal of Political Economy, 128(6):14101455.

Polachek SW (2017). Heterogeneity in the Labor Market: Ability and Information Acquisition. Eastern Economic Journal, 43(3):377-390.
Pratt JW (1981). Concavity of the log likelihood. Journal of the American Statistical Association, 76(373):103-106.
Prescott EC \& Visscher M (1980). Organization capital. Journal of Political Economy, 88(3):446-461.
Prokhorov A \& Schmidt P (2009). Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas. Journal of Econometrics, 153(1):93-104.
Qu F, Horrace CH \& Wu GL (2015). Wrong skewness and finite sample correction in parametric stochastic frontier models. Manuscript.
Rajapaksa DP (2015). Floods and property values: A hedonic property and efficiency analysis. Doctoral dissertation. Queensland University of Technology.
Reed WJ (2006). The normal-Laplace distribution and its relatives. In Balakrishnan N, Castillo E \& Sarabia JM (eds), Advances in Distribution Theory, Order Statistics, and Inference, pp. 61-74. Birkhäuser Boston.
Rho S \& Schmidt P (2015). Are all firms inefficient? Journal of Productivity Analysis, 43(3):327-349.
Richmond J (1974). Estimating the efficiency of production. International economic review, 15(2):515-521.
Ritter C \& Simar L (1997). Pitfalls of normal-gamma stochastic frontier models. Journal of Productivity Analysis, 8(2):167-182.
Romer PM (2015). Mathiness in the theory of economic growth. American Economic Review, 105(5):8993.

Roth AE (1979). Axiomatic models of bargaining. Springer-Verlag. Berlin.
Schmidt KD (2014). On inequalities for moments and the covariance of monotone functions. Insurance: Mathematics and Economics, 55:91-95.
Schmidt P \& Sickles RC (1984). Production frontiers and panel data. Journal of Business E Economic Statistics, 2(4):367-374.
Serfling RJ (1980). Approximation theorems of mathematical statistics. John Wiley \& Sons. New York.
Shabanpour H, Yousefi S \& Saen RF (2017). Future Planning for Benchmarking and Ranking Sustainable Suppliers Using Goal Programming and Robust Double Frontiers DEA. Transportation Research, 50D:129-143.
Shapiro C \& Stiglitz JE (1984). Equilibrium unemployment as a worker discipline device. The American Economic Review, 74(3):433-444.
Sharif NR \& Dar AA (2007). An Empirical Investigation of the Impact of Imperfect Information on Wages in Canada. Review of Applied Economics, 3(1-2):137-155.
Shepherd RW (1953). Cost and production functions. Princeton University Press. Princeton.
Shepherd RW (1970). The Theory of Cost and Production functions. Princeton University Press. Princeton.
Simar L \& Wilson PW (2009). Inferences from cross-sectional, stochastic frontier models. Econometric Reviews, 29(1):62-98.

Sklar A (1959). Fonctions de répartition à n dimensions et leurs marges," Publications de l' Institut de Statistique de l' Université de Paris, 8:229-231.
Sklar A (1973). Random variables, joint distribution functions, and copulas. Kybernetika, 9(6):449-460.
Smith MD (2008). Stochastic frontier models with dependent error components. The Econometrics Journal, 11(1):172-192.
Stevenson RE (1980). Likelihood Functions for Generalized Stochastic Frontier Estimation. Journal of Econometrics, 13:57-66.
Stirzaker D (2003). Elementary probability, 2nd ed. Cambridge University Press. Cambridge.
Styan GP (1973). Hadamard products and multivariate statistical analysis. Linear Algebra and its Applications, 6:217-240.
Tao J, Xin B \& Lian Y (2014). Measurement of Listed Company's Bargaining Power in the Credit Market: based on Two-tier Stochastic Model. Review of Investment Studies (04). http://en.cnki.com.cn/Article_en/CJFDTOTAL-TZYJ201408006.htm
Tauer L \& Fried H (2018). Wine hedonic pricing using the two-tier stochastic model. Paper presented at the Ithaca 2018 conference. https://www.wine-economics.org/2018-ithaca/2018-ithaca-scientific-program.
Tomini S, Groot W \& Pavlova M (2012). Paying informally in the Albanian health care sector: a twotiered stochastic frontier model. The European Journal of Health Economics, 13:777-788.
Torii A (1992). Technical efficiency in Japanese Industries. In: Caves RE, ed. (1992). Industrial Efficiency in Six Nations. MIT Press, 31-119.
Tran KC \& Tsionas EG (2015). Endogeneity in stochastic frontier models: Copula approach without external instruments. Economics Letters, 133:85-88.
Triebs TP \& Kumbhakar SC (2013). Productivity with general indices of management and technical change. Economics Letters 120(1):18-22.
Triebs TP \& Kumbhakar SC (2018). Management in production: from unobserved to observed. Journal of Productivity Analysis, 49(2-3):111-121.
Trivedi PK \& Zimmer DM (2005). Copula modeling: an introduction for practitioners. Foundations and Trends in Econometrics, 1:1.
Tsionas EG (2012). Maximum likelihood estimation of stochastic frontier models by the Fourier transform. Journal of Econometrics 170: 234-248.
Tsionas EG (2015). The role of management in production: theoretical and statistical implications. Manuscript. Lancaster University Management School.
Verbeek M (2012). A guide to modern econometrics, 4th ed. John Wiley \& Sons. Chichester, West Sussex.
Waldman DM (1982). A stationary point for the stochastic frontier likelihood. Journal of Econometrics, 18(2):275-279.
Waldman DM (1984). Properties of technical efficiency estimators in the stochastic frontier model. Journal of Econometrics, 25(3):353-364.
Wang PY (2017). A Six Component Panel Stochastic Model. Master Thesis. National Taiwan University.
Wang Y (2016a). Bargaining matters: an analysis of bilateral aid to developing countries. Journal of International Relations and Development. https://doi.org/10.1057/jird.2016.8
Wang Y (2016b). The effect of bargaining on US economic aid. International Interactions, 42(3):479-502.
Wei Z (2015). Financial Constraint, Government R\&D Subsidies and Corporate R\&D Investment: An Empirical study of China's Strategic Emerging Industries. Contemporary Finance E Economics, 11.
Wen HX, Liu XJ, Wang H \& Caputo Y (2016). The Influence of Financial Constraint and Agency Cost on Investment Inefficiency of GEM Listed Firms-Based on Two-Tier Stochastic Frontier Model. Journal of Mechanical Engineering Research and Developments, 39(2):584-592. DOI: 10.7508/jmerd.2016.02.038.
White H. \& Domowitz I (1984). Nonlinear regression with dependent observations. Econometrica, 52(1):143-161.
Withers CS (1981). Central limit theorems for dependent variables. I. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 57(4):509-534.

Wolff FC (2016). Bargaining Powers of Buyers and Sellers on the Online Diamond Market: A Double Perspective Non-Parametric Analysis. Annals of Operations Research, 244(2):697-718.
Woodward J (1980). Industrial Organization: Theory and Practice, 2nd ed. Oxford University Press. Oxford.
Xie C \& Li L (2018). The empirical test on investment efficiency and influence of equity incentive in supply-side structural reform: Based on the two-tier stochastic frontier approach. Applied Stochastic Models in Business and Industry, https://doi.org/10.1002/asmb.2304.
Xu H, Wang H, Zhou C \& Geng Z (2016). Dual Attributes of Government Intervention and China's Real Estate Prices -Is Chinese Government a Promoter of the Real Estate Prices? Accounting and Finance Research, 5(3):29-36.
Yan X \& Qi C (2017). Asymmetric information, bargaining power and the formation of Chinese export fruit price: based on the two-tier stochastic frontier model. Stat Inf Forum, 32:46-54.
Yaron D (1960). Resource allocation for dairy and field crops in the Negev area of Israel. PhD thesis. Iowa State University.
Yu L \& Liang T (2012). The performance of dividend policy-Based on financing constraints and agency cost trade-off. In Proceedings of 2012 International Conference on Information Management, Innovation Management and Industrial Engineering (ICIII), 1:455-459.
Zhang H, Zhang J, Yang Y \& Zhou Q (2017). Bargaining Power in Tourist Shopping. Journal of Travel Research, https://doi.org/10.1177/0047287517724917.
Zhang X \& Sun G (2015). Dual Characteristics of Exchange Rate and the Dispute on RMB Between China and the US-Is China a Currency Manipulator? Finance E Trade Economics, 8.
Zhang Z \& Zheng Z (2012). The Influence of Financial constraint and Agency Cost on Investment Inefficiency of listed Firms: An empirical Measurement based on Two-tier stochastic frontier Model Journal of Industrial Engineering and Engineering Management, 2:119-126.
Zheng Z \& Zhang Z (2012a) IPO Bookbuilding and Issue: Underpricing Effect or Overpricing Effect? An Empirical Measure based on Two-tier stochastic Frontier Model. Systems Engineering (03). http://en.cnki.com.cn/Article_en/CJFDTOTAL-GCXT201203004.htm
Zheng Z \& Zhang Z (2012b). Cause Decomposition and Re-examination of Chinese extreme IPO Returns-Evidence from Chinese Hybrid Verification-Bookbuilding Method Offerings. Journal of Shanxi Finance and Economics University, 04:37-47.
Zhengyan L \& Chuanrong L (1996). Limit theory for mixing dependent random variables. Kluwer Academic Publishers. Dordrecht.


[^0]:    ${ }^{1}$ The other is Data Envelopment Analysis that uses nonparametric mathematical programming, while SFA uses mostly parametric econometric techniques.
    ${ }^{2}$ This can be seen to be a compactification of a famous maxim that goes "Management is doing things right. Leadership is doing the right things". See chapter 6.

[^1]:    ${ }^{1}$ In chapter 2 we point to certain weaknesses of the authors' construction from a conceptual point of view that reveal their model to be in reality a single-tier SF model only. In chapter 5, we develop a structural model that leads validly to a 2TSF model in Nash bargaining situations.

[^2]:    ${ }^{2}$ We were not able to obtain full English copies for some of the papers from China referenced here, so for them we rely on the available abstracts.

[^3]:    ${ }^{3}$ And, as luck would have it, it contains typographical errors in two formulas in the main text (eq. 11 and 12). The corresponding formulas in their Appendix (eq. A. 10 and A.13) are the correct ones and should be used instead.
    ${ }^{4}$ This paper is part of the research undertaken for the present PhD thesis, and its results appear in chapter 3.

[^4]:    ${ }^{5}$ We provide the correct composite error density for this case in an upcoming contribution to the collective volume Sickles R \& Parmeter C (eds), "Advances in Efficiency and Productivity Analysis".
    ${ }^{6}$ I would like to thank professor Kristiaan Kerstens for bringing this literature to my attention.

[^5]:    ${ }^{1}$ Their exact words were "the demand curve is traditionally defined as the maximum quantity of labor demanded at any wage level".

[^6]:    ${ }^{2}$ I would like to thank professor Subal Kumbhakar for bringing my attention to these matters.

[^7]:    ${ }^{3}$ This is a very different situation than Pissarides' stochastic model in chapter 6 of his book, where productivity uncertainty exists only ex ante, and it is resolved once the firm and the worker meet, prior to fill the position.

[^8]:    ${ }^{1}$ In Torii (1992) section 2.3.1, specific structural models of various sources of inefficiency in production are developed that lead to a Half-normal distribution for the one-sided component in a single-tier SF production model. It is the only research that we know of where the wish of professor Greene (1985) is granted, the wish being (quote) "It would be refreshing to see some space devoted to ah attempt to arrive at a distribution for the 'stochastic inefficiency' through a behavioral analysis of the firm."

[^9]:    ${ }^{2}$ Stevenson (1980) commented early on in the development of SFA that such an assumption may not always hold, even in the single-tier SF framework. We will have to say more on this later in the chapter.
    ${ }^{3}$ And exactly as it happened with the Half-normal single-tier SF specification and the Skew-normal family of distributions, a statistician later came and baptized the 2TSF specification as a distribution family, the "Normal-Laplace", see Reed (2006).

[^10]:    ${ }^{4}$ We note again that one should use the mathematical expressions that are found in the Appendix of that paper, because formulas (11) and (12) in its main text have typographic errors.

[^11]:    ${ }^{5}$ Part of the material in this section has been published in Papadopoulos (2015a). In that paper we used a different notation, but here we revert to the standard notation used in the literature.

    ## 

[^12]:    ${ }^{6}$ These densities were used in the field of stochastic frontier analysis years before being baptized "Skew-normals" and formally studied as a distinct distribution family by A. Azzalini.

[^13]:    ${ }^{7}$ We must note that the more mainstream terminology is that the "Corrected" OLS (COLS) method refers to the case where the intercept and the OLS residuals are shifted by the use of the maximum OLS residual obtained, while the case where we shift by the estimated expected value of the error term is called "Modified" OLS (MOLS) labels. Kumbhakar, Parmeter and Zelenyuk (2018) footnote 15, lay out the history of the use of these two terms in the literature and provide a convincing argument as to why we should swap labels, and we follow them here.

[^14]:    ${ }^{8}$ We note that one weakness of the COLS/MM estimator is that it allows for a limited modeling of conditional heteroskedasticity of the one-sided error components, i.e. for the use of covariates as conditioning arguments for the moments/measures of inefficiency (no more than one covariate per unknown parameter). Moreover, in such a case the asymptotics of the estimator must be worked anew, since now the convenient separation of unknown parameters from the data no longer holds.s ${ }^{s}$.

[^15]:    ${ }^{9}$ Superficially the operation is the opposite from the one usually applied for the COLS correction for single-tier SF models, but in our 2TSF case this is the correct operation.

[^16]:    ${ }^{10}$ See eg. Kendall (1945) ch. 3.

[^17]:    ${ }^{11}$ The kapa-statistics and the MM estimator are general tools, and can be used with any error distributional specification that contains no more than four unknown parameters.

[^18]:    ${ }^{12}$ We generally follow here Cameron and Trivedi (2005), the relevant parts of chapters 5 and 6.

[^19]:    ${ }^{13}$ I would like to thank professor Stylianos Arvanitis for his guidance related to the results in this section.

[^20]:    ${ }^{14}$ We adopt here the terminology according to which "asymptotic" refers to the large-sample finite approximation.

[^21]:    ${ }^{15}$ The reader is referred to the original paper for more detailed information on the data sample.

[^22]:    ${ }^{16}$ For shape parameter smaller than unity, the density of the Gamma distribution becomes log-convex, see Bergstrom \& Bagnoli (2005).

[^23]:    ${ }^{17}$ The Kumaraswamy distribution is a "cousin" to the Beta distribution, and was introduced in the field of Hydrology by Kumaraswamy (1980). A comprehensive exposition closer to home is Jones (2009). When one of the parameters of the distribution equals unity, it coincides with the Beta distribution with same parameter values.

[^24]:    ${ }^{1}$ Material from this section has been included in a joint paper with Subal Kumbhakar and Chris Parmeter that is currently in its writing stage.

[^25]:    ${ }^{2}$ In Feund's paper there is an obvious typo in the expression for the MGF, specifically in eq. 2.2 and 2.3.

[^26]:    ${ }^{3}$ We note that this issue arises only because we ultimately use the difference $w-u$ of two variables following Freund's distribution. If we had data on the two variables, all four parameters would be identifiable.

[^27]:    ${ }^{4}$ One could ask why this computation cannot be used also to examine the existence of statistical dependence between the two one-sided error components. There are two reasons why it should not. First, it would be a point estimate, usually with a value rather close to zero, and without the variance of the expression, we will not be able to formulate a formal statistical test. Second, as we have already said the specification allows also for pure non-linear dependence with zero covariance.

[^28]:    ${ }^{5}$ The theorem appeared originally in a 3-page note, Sklar (1959). See also Nelsen (2006) p. 21, and Sklar (1973) for an elaboration and various foundational results.

[^29]:    ${ }^{6}$ For a different method to handle discrete random variables in Copula modeling, see Danaher \& \& Smith (2011).

[^30]:    7 Although we must mention that even this result does not cover coefficients of a regression, only location parameters of the distributions involved.
    8 The Student's Copula is also a member of the elliptical class and it is preferred in Finance applications because it allows for extreme-values ("tail") dependence, something that the Gaussian Copula lacks, this being perhaps its most serious weakness.

[^31]:    ${ }^{9}$ See also Mari \& Kotz (2001), p 155.

[^32]:    ${ }^{1}$ See Mas-Colell, Whinston \& Green 1995, pp 838-846, for a compact introduction.

[^33]:    ${ }_{2}$ Nash's papers contained mostly verbal arguments that may not appear fully transparent to the unprepared reader. Detailed analysis and explicit mathematical derivations of the various properties and results can be found in Roth (1979).

[^34]:    ${ }^{3}$ An extension, as suggested by Binmore et al. (1986), would be to assume that $s_{0}(f)$ is strictly positive, being the gain for the firm if it were to collaborate with some other worker. Although this does not affect the solution to our model, it becomes relevant for the interpretation of empirical results, something we discuss later.

[^35]:    ${ }^{4}$ See for example Lombera (2007), where a tendency to not even negotiate from the part of the employees but accept the first firm's offer is documented.
    ${ }^{5}$ The variables in [5.11] can also be in logarithmic form, which is consistent with the usual log-log or semi-log econometric specifications for the wage equation.

[^36]:    ${ }^{6}$ This is true both when the situation is one of a prospective new hire, but also when an existing employee negotiates her wage anew.

[^37]:    7 This and the corresponding possible constraints for the worker side are specific examples of how asymmetries can be incorporated already in the surplus functions, as Binmore et al. (1986) have stressed.

[^38]:    8 "Unskilled" in the sense of no formal educational background.

[^39]:    ${ }^{9}$ In fact, there are cultures where after a successful negotiation the parties are expected to "look displeased" as a signal to the other party that they have indeed foregone something valuable to them during the negotiation.

[^40]:    ${ }^{10}$ Experts on negotiation theory will point out that such tactics belong to the "positional bargaining" strategy, and then set out to explain why it is an inefficient way to bargain.

[^41]:    ${ }^{11}$ The data was collected as part of the Regional Program on Enterprise Development (RPED), organized by the World Bank. The Ghana surveys were conducted by a team from the Centre for the Study of African Economics at the University of Oxford and from the Department of Economics, University of Ghana at Legon. The project received support from the Swedish, Norwegian, United Kingdom, Canadian, and Dutch governments. The full data set is available from the web site of the Centre for the Study of African Economies at the University of Oxford (CSAE): http://www.csae.ox.ac.uk.

[^42]:    ${ }^{12}$ The tests were performed using the on-line web tool "MVN: a web-tool for assessing multivariate normality (ver. 1.6)" in http://www.biosoft.hacettepe.edu.tr/MVN/. The tool uses the R-package "MVN" and was developed by Korkmaz, Goksuluk \& Zararsiz (2014), where one can find details about the MVN tests implemented.

[^43]:    ${ }^{1}$ It is usually attributed to Peter Drucker, something that remains unverified and occasionally disputed. What Drucker (1986) did wrote is "without productivity measurements, (the business) does not have control" (p. 83) and "work implies not only that somebody is supposed to do the job, but also accountability, a deadline, and finally the measurement of results, that is, feedback from results on the work and on the planning process itself." (p. 94), quotes that essentially say the same thing. The interested reader should also keep in mind the distinction between management and leadership. Leadership tends to be allergic to measurement, since, in order to achieve the expansion of existing boundaries, it tends to project a sense of "boundlessness".

[^44]:    ${ }_{2}$ An inspired example of high-quality logical speculation can be found in Kaldor (1934), where the author decomposed management into "supervision" and "co-ordination" (that roughly correspond to the distinction between management and leadership as we use the terms here) and argued that the latter is in fixed supply for each firm, even in the long-run, rationalizing in this way why firms are ultimately constrained as regards their size (due to the existence of an ever-fixed factor of production).

[^45]:    ${ }^{3}$ Covey (1989) p. 101, attributes the quote to W. Bennis and P. Drucker. Bennis \& Nanus (1985) do write (p. 21), "Managers are people who do things right and leaders are people who do the right thing."

[^46]:    ${ }^{4}$ It was later revealed in Bloom, Genakos, Sadun \& Van Reenen (2012) that the consulting firm was McKinsey.
    ${ }^{5}$ In later papers "operations" has been subsumed as a category to "monitoring".

[^47]:    ${ }^{6}$ In a personal communication in December 2017, prof. Bloom increased the current estimate to USD 500.
    ${ }^{7}$ This comes at a methodological price: no more telephone interviews, but self-reporting, which does not permit to control for biases in the same degree.

[^48]:    ${ }^{8}$ In Yaron (1960) we find also the following remark (p.66): "Hardcopf has suggested that the residuals from an empirically estimated production function may be considered as representing the share of the management factor in output and an error term", the reference being to an unpublished MS thesis by R.W. Hardcopf in 1956 at Iowa State University. I take here the opportunity to thank Jeffrey Kushkowski, Business and Economics Librarian at the Iowa State University, for his assistance regarding obscure documents in his library.

[^49]:    ${ }^{9}$ This is the peril of "mathiness" as conceived in Romer (2015), where mathematical symbols do not faithfully represent what they are supposed to represent, resulting in a mathematical model that does not really represent the purported situation under study, leading to apparently rigorous but essentially irrelevant conclusions.
    ${ }^{10}$ See Bloom \& Van Reenen $(2007,2010)$ and Bloom et al. (2014) for discussions about the heterogeneity of management practices across firms.

[^50]:    ${ }^{11}$ Whether KPI's are used by a business and how many of them are used, is an aspect to contribute in measuring the quantity of management.

[^51]:    ${ }^{12}$ A known Chinese proverb asserts that "He who rides a tiger is afraid to dismount". While it is usually interpreted in relation to the undertaking of "dangerous" activities, it is also a great image for the risky world of entrepreneurial activity, and how success stories can happen due to forces beyond control.
    ${ }^{13}$ Bloom et al. (2016), in one of their panel data regressions, do indeed apply a Fixed Effects model to separate management from unexplained heterogeneity. See also Colombi, Kumbhakar, Martini \& Vittadini (2014) for a comprehensive approach to measure firm heterogeneity using the Closed Skew Normal family of distributions.

[^52]:    ${ }^{14} 150,000$ such, as the consulting firm running the program declares on its website.

[^53]:    ${ }^{15}$ Technology as a recipe is not a new metaphor, see for example Kerstens, O'Donnell \& Van de Woestyne (2015), p. 5.
    ${ }^{16}$ Confusing use of terms and concepts is not rare in the literature surrounding organization and management. Prescott \& Visscher (1980) titled their paper "Organizational Capital". It turns out that the authors present three different models under this label, the first two dealing with information accumulation for better matching of employees and positions, and the third relating to human capital. Certainly, matching successfully an employee's skill set with a position is important for efficiency, but it is hardly more than the bare minimum a management system must accomplish. The real challenge of management is the consistent and successful collaboration and combination of inputs, after they have been correctly positioned. Atkeson \& Kehoe (2005) use the term "organization capital" with the meaning "accumulated plant-specific knowledge".
    ${ }^{17}$ See for example Gibbons and Henderson (2013), where the situation is formalized through the concept of "relational contracts" and repeated "trust" games.
    ${ }^{18}$ See for example Woodward (1980) ch. 12.

[^54]:    ${ }^{19}$ The Leontief production function is a special case where fixed input ratios and zero marginal products are hard-wired in the structure of the model, and so the concept of "increasing one input while keeping the others fixed" is inapplicable.

[^55]:    ${ }^{20}$ This contrast possibly implies that modeling the management contribution in cost-efficiency models would require a different modeling approach, but this is outside our current scope of research.

    40

[^56]:    ${ }^{21}$ This result does not depend on the chosen exponential base for $h(\cdot)$.
    ${ }^{22}$ Tsionas (2015) analyzes in detail a profit-maximizing model with management, and arrives at the same result, namely, that economic optimization will result in a level of management below its technically optimal level, if one exists, or in general in a level that will leave some technical efficiency opportunities unrealized.
    ${ }^{23}$ The profit-maximizing problem for a price-taking firm is more demanding. We show in the Technical Appendix that the second-order conditions are satisfied if $F(\mathbf{x})$ is jointly strictly concave and $h^{\prime \prime}(m)+\left(h^{\prime}(m)\right)^{2}<0$.

[^57]:    ${ }^{24}$ But see the next section about the existence of internal inefficiency and how it can reverse the observed correlations between management and conventional inputs.

[^58]:    ${ }^{25}$ With hindsight, we find Hall and Winsten (1959) a fascinating read, seeing how they focus intensely on the issue of "different environments" and stressing that they must be taken into account in order to assess accurately the efficiency of firms and managers. They don't neglect to criticize Farrell (1957) for not analyzing "the crucial concept of environment".

[^59]:    ${ }^{26}$ Compare this with the approach of Alvarez, Arias and Greene (2005), where as mentioned earlier, through the specification of the technically optimal level of management they obtain the inefficiency component as a function of the deviation of management from its optimal level.
    ${ }^{27}$ See Shapiro \& Stiglitz (1984) for one of the original formulations.

[^60]:    ${ }^{28}$ Easily accessible formal proofs of this known and intuitive result appeared only relatively recently; see Schmidt K (2014) and Egozcue (2015).

[^61]:    ${ }^{29}$ Endogeneity due to simultaneity bias and correlation of inputs with the error term are in any case long recognized as issues present in production function specifications, see for example Olley and Pakes (1996).

[^62]:    30 The full data set is "Manufacturing: 2004-2010 combined survey data (AMP)", freely available at http://worldmanagementsurvey.org/survey-data/download-data/download-survey-data/.

[^63]:    ${ }^{31} \mathrm{http}: / / \mathrm{www} . b i o s o f t . h a c e t t e p e . e d u . t r / M V N /$. The reference is Korkmaz, Goksuluk \& Zararsiz (2014).

[^64]:    ${ }^{32}$ By estimating the model without a copula density, we established that this is a consequence of the inclusion of the Copula.

[^65]:    ${ }^{33}$ But others do. Statistically, we have a true parameter at the boundary, and a maximum likelihood estimator with a singular covariance matrix. These affect the asymptotic properties of the estimator, but this is not our concern here. The easy solution is to re-estimate the model using the density of $\varepsilon=w-u$ only.

[^66]:    ${ }^{34}$ We used the Freedman-Diaconis (1981) "(1.8) rule" to determine optimal bin-length $=2 \mathrm{IQR} / n^{1 / 3)}$.

[^67]:    ${ }^{35} \mathrm{He}$ nevertheless points out that the sample skweness is only suggestive, and sensitive to the assumption that the symmetric random disturbance $v$ is normal: under non-normality the population skewness could be positive even in the presence of a one-sided inefficiency component.
    ${ }^{36}$ The composite error term in the single-tier SF Half-normal specification follows a Skew-normal distribution and the intricacies of maximum likelihood estimation when the true value of the skewness (or "slant") parameter is zero, have also been investigated in the relevant statistical literature (see Azzalini and Capitanio 2014, ch. 3, for an exposition).

[^68]:    ${ }^{37}$ The data set is provided as part of Verbeek (2012) and can be freely downloaded from the textbook's "student companion" website. The file is "LABOUR2".

[^69]:    ${ }^{38}$ Since this is the 2nd time that this happens in the chapter, the reader would be justified to wonder whether it is some artificial result imposed by the properties of the statistical specification. The result did not appear in other samples, but we reserve judgment since the 2TSF Generalized Exponential specification is a newborn.

