The University of Manchester

Economics
Discussion Paper Series
EDP-0812

On the Production of Economic Bads and the Estimation of Production Effciency when the Dependent Variable is a Count

Eduardo Fé-Rodríguez

December 2008

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

On the production of Economic Bads and the Estimation of Production Efficiency when the Dependent Variable is a Count

Eduardo Fé-Rodríguez*
e.fe-rodriguez@manchester.ac.uk

Health Economics Research at Manchester
University of Manchester
Oxford Road
Manchester
M13 9PL

Incarnation 15 November 2008

Abstract

This article develops a (cost) stochastic frontier model suitable for dependent variables measured in discrete amounts. The model is put into context by developing a suitable theoretical framework and a family of econometric models is defined to estimate this type of production technologies. The performance of the method in small samples is evaluated via Monte Carlo experiments and its usage is illustrated via an application.

Key Words: Stochastic Frontier, Count Data, Discrete Convolution, Delaporte Distribution.

JEL: **C01, C13, C25, C51**.

*The author the comments and encouragement received from William Greene, Antonio Alvarez, Mehdi Farsi, Matt Sutton, Stephen Birch, Bruce Hollingsworth, and participants in NAWPA V, ECHE-Rome 2008, EWEPA X, the Meeting of the Swiss Society of Economics and Statistics and Econometrics Seminar at Oviedo University for helpful comments

1 Introduction

This article introduces a model of stochastic frontiers applicable to those situation when the dependent variable is an economic bad measured in discrete (and non-negative) amounts. The need for such a model arose in a study of infant mortality in England aimed at detecting what areas in the country were efficient in controlling the levels of infant deaths. The dependent variable of interest is this study was an economic bad (a commodity with negative marginal utility), it took values on the set of natural numbers, and the observed count was typically small.

Existing Stochastic Frontier Methods with roots in the seminal article by Aigner, Lovell and Schmidt [Aigner et al., 1977] (ALS in what follows) are feasible, but not optimal, devices for efficiency analysis in this context. From an econometric perspective, these methods have been developed with continuous random variables in mind, and therefore they would ignore the discrete nature of the data, which will result in at best inefficient estimates. From a technical perspective, it is known that associated to ALS methods, there is *distance function* (REF) relating output, Y, inefficiency, $D \geq 1$ (for economic bads) and the frontier of production possibilities, Q, via a multiplicative scheme, so that $Y = Q \times D$. However, if output and inefficiency are discrete valued, a multiplicative scheme such as this will rarely solve the equality $Y = Q \times D$ -for example, if $Y = 7$ (a prime number) only the trivial solution is admissible. In view of these problems, this article proposes a theoretical environment leading to a suitable *distance function* for the study of the production of bads measured in discrete amounts. From this environment a class of econometric models is derived in order to implement of the theoretical distance measure.

The next section is devoted to the development of the theoretical framework. The analysis in this section borrows from Shephard [Shephard, 1970], Fare, Grosskopf and Lovell [Fare et al., 1994], [Kumbhakar & Knox Lovell, 2003] to present an axiomatic framework describing the production of economic bads. The key features of the theoretical model are minimization of outputs and *lack of control* as drivers of the production of bads. The former feature seems to be evident given the negative marginal utility of economic bads, while the later issue relates to observed positive outputs even when no inputs are present in the production scheme (for example, infant deaths will be observed even when there are no risk factors); agents may try to control the amount of economic bad produced by a given set of inputs, however nothing grants that control be effective, and therefore the potential loss of utility caused by a fixed level of input may be unbounded. This discussion will lead us to analyse the role of 'discreteness' in the measurement of efficiency and to the introduction of a new additive distance-function adapted to the new environment.

Section 3 proposes an econometric equivalent of the distance function devised in Section 2. The resulting econometric model is a convolution of the two discrete probability distribution representing the stochastic frontier and the stochastic inefficiency levels. The basic model does not specify a particular probability distribution for the random variables involved, but it is argued that models based on convolutions of Mixed Poisson ($\mathcal{MP}$) and standard Poisson ($\mathcal{P}$) random variables arise naturally. Like ALS these models may be estimated by Maximum Likelihood, but unlike with ALS and other SFM, they are capable of separating unobserved heterogeneity from inefficiency. This happens through the inclusion of the mixing parameter in the $\mathcal{MP}$ part of the model (see [Hausman et al., 1984] or [Karlis & Xekalaki, 2005]). Among the class of convolutions of this type, the Delaporte family of models (Delaporte, [Delaporte, 1962], Ruohonen [Ruohonen, 1989] and Willmot and Sundt [Willmot & Sundt, 1989]) appears to be a *normal* choice, and this model is studied at length in the article.

As with likelihood based SFM, the ones presented here may suffer of a problem of *near identification* leading to a loss of precision of the estimates. Riter and Simar [Ritter & Simar, 1997] and [Bandyopadhyay & Das, 2008] discuss and characterize how, under some limit conditions, the models degenerate into the convolution of two equally distributed random variables, with the consequent loss of identification in some of the parameters of the models. Monte Carlo evidence seems to suggest that the issue of near identification may arise in some extreme cases, but if it does, the models themselves will provide indications of the potential loss of precision (via the estimated value of the parameters driving unobserved heterogeneity), so that empirical applications of the models may be safely undertaken.

Section 4 illustrates the application of the Delaporte model to the original problem of studying efficiency in the production of infant deaths in England. The study sheds light on the role of social, economic and environment factors to explain infant mortality, and at the same time, it allows further study of the issue of near identification in practice. Section 5 gives some closing remarks.

2 The Production Technology of Economic Bads

This section formalizes the axiomatic framework within which the production of economic bads takes place. Convex Duality Theory (see, Blackorby et al ([Blackorby et al., 1978])) drives the discussion toward the definition of the production function of economic bads and the construction of efficiency measure functions, which are the ultimate goals of this section. The general theoretical framework presented below is valid regardless of whether outputs are measured in discrete or continuous amounts. However, when it comes to study efficiency in production, it becomes necessary to distinguish situations where outputs are measured in discrete amounts from those cases where outputs are measured in continuous amounts: multiplicative schemes arise naturally in the construction of *distance functions* for measuring efficiency in the continuous case; however these schemes will not be able to accommodate production measured in discrete quantities.

2.1 Technology Set and Production Function

The starting point of the analysis is the definition of the Technology Set, which describes all the feasible input-output combinations. Let $\mathbf{x} \in \mathbb{R}^{K+}$ denote a set of inputs which combine to bring forth a single economic bad in quantities $y \in \mathbb{R}^{+}$, where $\mathbb{R}^{+}$ is the set of non-negative real numbers, and $\mathbb{R}^{K} = \prod_{k=1}^{K} \mathbb{R}$. The Technology Set has been traditionally defined as the set $\mathcal{S} := \{(\mathbf{x}, y) : \mathbf{x} \text{ produces } y\}$. The set $\mathcal{S}$ satisfies the following restrictions.

Proposition 1 *The technology set $\mathcal{S}$ satisfies the following assumptions:*

- $\mathcal{S}$*.1* *For every $(\mathbf{x}, y) \in \mathcal{S}$, (i) $(c\mathbf{x}, y) \in \mathcal{S}$ whenever $0 \leq c \leq 1$ and (ii) $(\mathbf{x}, cy) \in \mathcal{S}$ whenever $1 \leq c$.*
- $\mathcal{S}$*.2* *For all y, (i) $(\mathbf{0}; y) \in \mathcal{S}$ and (ii) if $(\mathbf{x}; 0) \in \mathcal{S}$ then $\mathbf{x} = \mathbf{0}$*
- $\mathcal{S}$*.3* *Let $(\mathbf{x}, y) \in \mathcal{S}$ and define $y^* = y^*(\mathbf{x}) = \inf_y \{(\mathbf{x}, y) \in \mathcal{S}\}$. Then $(\mathbf{x}, y^*(\mathbf{x})) \in \mathcal{S}$ for all $\mathbf{x}$. Similarly, let $\mathbf{x}^* = \mathbf{x}^*(\mathbf{x}, \mathbf{y}) = \mathbf{x} \sup_{\lambda \geq 1} \{(\lambda \mathbf{x}, y) \in \mathcal{S}\}$. Then, $(\mathbf{x}^*(\mathbf{x}, \mathbf{y}), y) \in \mathcal{S}$*

There are two salient features implicit in the above definition. Firstly the output variable is a bad and its incidence needs to be minimized. Secondly, there is an inherent lack of control in this productive process: even in the absence of inputs, non-negative (and unbounded) outputs are feasible. To see this note that the monotonicity condition $\mathcal{S}.1$ implies that if $\mathbf{x}$ can produce y harm, then it also may cause harm beyond that level; on the other hand, it is not possible to reduce the levels of harm without reducing the magnitude of risk factors. As it happens in the traditional framework, this assumption forces increases in inputs to deliver increases in output, thus ruling out, in principle, inputs with a negative marginal product. Condition, on the other hand, $\mathcal{S}.2$ implies that negative bads are producible without any need of inputs -therefore, risk factors may be absent, and yet economic bads may be observed -in fact, $\mathcal{S}$ is not bounded above. Together with $\mathcal{S}.1$, assumption $\mathcal{S}.2.i$ formalizes what we meant by lack of control in the production of bads. Condition $\mathcal{S}.2.ii$ assumes that zero-incidence of a bad can only happen if the inputs are absent in production. This latter condition may be assumed without loss of generality, because if there was a value $\dot{y}$ such that $y \geq \dot{y}$ for every $(\mathbf{x}, y)$, then $\mathcal{S}$ could be re centered, and defined as $\mathcal{S} := \{(\mathbf{x}, y - \dot{y}) : \mathbf{x} \text{ produces } y\}$. Condition $\mathcal{S}.3$ ensures that we will be able to define a production function. This condition defines the set of cluster points of $\mathcal{S}$, and it establish a lower bound for outputs. Note that convexity is not among the assumptions introduced above so that outputs could be measured in discrete amounts. It also allows a rich variety of functional forms for $y^*(\mathbf{x})$ provided that the latter function is monotonically non-decreasing (as required by condition $\mathcal{S}.1.i$). This latter fact will be relevant when defining the production function.

Associated to $\mathcal{S}$, there are an input and output sets defined, respectively as $\mathcal{X}(y) := \{\mathbf{x} : (\mathbf{x}, y) \in \mathcal{S}\}$ and $\mathcal{Y}(\mathbf{x}) := \{y : (\mathbf{x}, y) \in \mathcal{S}\}$. Conditions $\mathcal{S}.1 - \mathcal{S}.3$ above carry implicit the properties satisfied by the input and output sets. In particular we have the following:

Theorem 2 *Let $\mathcal{X}(y)$ be defined as above, and let $\mathcal{S}.1 - \mathcal{S}.4$ hold. Then,*

$\mathcal{X}.1$ (i) $\mathcal{X}(\mathbf{0}) = \{0\}$ *and* (ii) $0 \in \mathcal{X}(y) =$ *for every* y

$\mathcal{X}.2$ $\mathcal{X}(y) \to \mathbb{R}^{K+}$ *as* $y \to \infty$

$\mathcal{X}.3$ $\mathbf{x} \in \mathcal{X}(y) \Rightarrow c\mathbf{x} \in \mathcal{X}(y)$, *whenever* $0 \leq c \leq 1$

$\mathcal{X}.4$ $\mathcal{X}(y)$ *is a closed set.*

$\mathcal{X}.5$ $\mathcal{X}(y)$ *is a convex set.*

Proof. $\mathcal{X}.1$, $\mathcal{X}.2$ and $\mathcal{X}.3$ follow directly from the definition of $\mathcal{S}$ and its properties. To prove $\mathcal{X}.4$, let y be arbitrary, and suppose that there exists and adherent point of $\mathcal{X}(y)$, namely $\mathbf{x}$, which is not in $\mathcal{X}(y)$. Then, for every ball $\mathcal{B}\left(\mathbf{x}; \frac{1}{n}\right)$ we can find a $\mathbf{x}_n = \lambda_n \mathbf{x}$ on the ray from the origin and through $\mathbf{x}$ such that, for all n, $\mathbf{x}_n$ produces y, $\mathbf{x}_n \geq \mathbf{x}_{n-1}$, and, since $\mathbf{x}$ does not produce y, $\mathbf{x}_n < \mathbf{x}$. The sequence $\{\mathbf{x}_n\}$ is bounded, monotone increasing and converges to $\mathbf{x}$, from which follows that $\mathbf{x} = \sup_n \{(\mathbf{x}_n, y) \in \mathcal{S}\}$. By $\mathcal{S}.3$ the point $(\mathbf{x}, y) \in \mathcal{S}$ so that we have meet a contradiction, and the proof follows. Finally, convexity is just a consequence of $\mathcal{X}.3$: if $\mathbf{x}$,$\mathbf{x}^{\star}$ are arbitrary in $\mathcal{X}(y)$ and $\theta \in (0,1)$ is also arbitrary, $\mathbf{z} = \theta\mathbf{x} + (1-\theta)\,\mathbf{x}^{\star}$ is such that $(\mathbf{z}, y) \in \mathcal{S}$ and so convexity follows. ■

Furthermore $\mathcal{S}.1 - \mathcal{S}.3$ define input set as the interval $\mathcal{Y}(\mathbf{x}) = [y(\mathbf{x}), \infty)$; as a consequence of this, it is immediate that

$\mathcal{Y}.1$ $\mathcal{Y}(\mathbf{0}) = \mathbb{R}^{+}$.

$\mathcal{Y}.2$ *For every* $\mathbf{x}$, $\mathcal{Y}(\mathbf{x})$ *is a closed set and bounded below -by 0.*

$\mathcal{Y}.3$ $\mathcal{Y}(\lambda\mathbf{x}) \subseteq \mathcal{Y}(\mathbf{x})$ *whenever* $\lambda \geq 1$.

$\mathcal{Y}.4$ *If* $y \in \mathcal{Y}(\mathbf{x})$ *then* $\lambda y \in \mathcal{Y}(\mathbf{x})$ *for* $\lambda \geq 1$.

The two sets just described will allow us to introduce the definition of the production function of economic bads in terms of the boundary of the output set for every input level.

Definition 3 *The production frontier of economics bads is the mapping*

$$F(\mathbf{x}) := \inf_{y} \{y \in \mathcal{Y}(\mathbf{x})\} = \inf_{y} \{\mathbf{x} \in \mathcal{X}(y)\} \tag{1}$$

In our single-output environment, the production function corresponds with to the output efficient subset of $\mathcal{Y}(\mathbf{x})$ (which includes those $y \in \mathcal{Y}(\mathbf{x})$ such that $y^{*} \notin \mathcal{Y}(\mathbf{x})$ whenever $y \geq y^{\star}$). Under the properties of the technology set $\mathcal{S}$, it is straightforward to establish that the production function is a mapping from $\mathbb{R}^{K+}$ into $\mathbb{R}^{+}$ such that that $F(\mathbf{0}) = 0$. $F(\lambda\mathbf{x}) \geq F(\mathbf{x})$ whenever $\lambda \geq 1$. It is quasi-convex function, because $\{\mathbf{x} : F(\mathbf{x}) \leq a\}$ is convex for every a -given the properties of the input set.

2.2 Efficiency Measures

The above definitions and theorems allow the definition of measures of efficiency in the production of economic bads, which are the key element of this article. The first definition is an Output-Oriented measure of technical efficiency, partly inspired in the work of Koopmans (1951), Debreu (1951) and Farrell (1957). In essence we present a convenient modification of what these authors call a *Distance Function*

Definition 4 *Consider the technology set* $\mathcal{S} \in \mathbb{R}^{K+} \times \mathbb{R}^{+}$. *An Output-Oriented*[1] *measure*

[1] An input oriented equivalent could be easily defined (in terms or maximal expansion of the input vector, given an output level), and we leave to the reader the formalization of the result.

of technical efficiency in the production of economic bads is a mapping

$$D(\mathbf{x}, y) = \inf_{\lambda} \{\lambda y \in \mathcal{Y}(\mathbf{x})\} \tag{2}$$

or equivalently

$$D(\mathbf{x}, y) = \inf_{\lambda} \{\lambda y \geq F(\mathbf{x})\} \tag{3}$$

The definition establishes that the above distance function is linear in y -this follows from equation (2)- so that there exits a function $D(\mathbf{x})$ such that $D(\mathbf{x}, y) = D(\mathbf{x})y$. Since, by definition,$d(\mathbf{x}, F(\mathbf{x})) = 1$, $D(\mathbf{x}) = 1/F(\mathbf{x})$. Therefore, a producer observes a quantity y of an economic bad, resulting from the equality

$$y = F(\mathbf{x}) \cdot D(\mathbf{x}, y), \tag{4}$$

The level of (output oriented) technical efficiency is thus $D(\mathbf{x}, y) \geq 1$ with equality only if the producer is *efficient*.

The above measure won't be suitable if the production of the bad of interest is measured in discrete amounts. This is due to the the multiplicative scheme underlying $D(.)$: even if $F(.)$ and $D(.)$ where restricted to be integers themselves, we may not be able to solve the equality $y = F(\mathbf{x}) D(\mathbf{x}, y)$ for $D(.)$. The solution is to use an additive scheme instead, since this naturally accommodates the discrete nature of the output variable. Let $Q(\mathbf{x})$, with $Q : \mathbb{R}^+ \mapsto \mathbb{Z}^+$, denote the production function of the discrete-measured bad. Then we have the following definition.

Definition 5 *Consider the technology set $\mathcal{S} \in \mathbb{R}^{K+} \times \mathbb{Z}^+$. An Output-Oriented measure of technical efficiency in the production of discrete economic bads is a mapping*

$$D^*(\mathbf{x}, y) = \min_{l \in \mathbb{N}} \{y - l \in \mathcal{Y}(\mathbf{x})\} \tag{5}$$

or equivalently

$$D^*(\mathbf{x}, y) = \min_{l \in \mathbb{N}} \{y - l \geq Q(\mathbf{x})\} \tag{6}$$

In the above definition $D^*(\mathbf{x}, y)$ is integer valued, and linear in y, so that for any m, $D^*(\mathbf{x}, y + m) = D^*(\mathbf{x}, y) + m$. Since $D^*(\mathbf{x}, Q(\mathbf{x})) = 0$ by definition, then for any non-negative integer y,

$$y = Q(\mathbf{x}) + D^*(\mathbf{x}, y). \tag{7}$$

This equality in $D()$ may be solved exactly for integer valued y, $Q(.)$ and $D(.)$.

3 Econometric Estimation

The focus of what follows is the empirical implementation of the efficiency measures defined in equations (3) and (6). Throughout, it is assumed that a random sample is available with observations of Y_i (the non-negative quantity of economic bad produced by the i^{th} cross-sectional unit) and $\mathbf{X}_i$ (a $k \times 1$ vector of inputs) for $i = 1, \ldots, n$. Y_i could be measured in discrete or continuous amounts. We shall denote by $(y_i, \mathbf{x}_i)$ an observed pair of values.

Estimation (3) and (6) in practice will require the definition of a parametric econometric model for the identities found in equations (4) and (7). The first element of each of these equations is the production function of Y_i, which depends on the contribution of a group of inputs. Let $\mathbf{X}_{1i}$ be the k_1 subset of elements of $\mathbf{X}_i$ collecting these inputs, and let $\beta_1 \in \mathcal{B}_1$ be a non-observable (although estimable) $k_1 \times 1$ vector summarizing the contribution of each element in $\mathbf{X}_{1i}$. It is useful to imagine that production takes place in two steps. Initially, the production inputs are transformed via the latent, deterministic mapping[2] $\rho : \mathbb{R}^{k_1} \times \mathcal{B}_1 \mapsto \mathbb{R}^+$. This mapping represents the expected output given the levels of input. Expected output will generally disagree with the actual output of the productive process, mostly due to the fact that productive units are subject to favorable and unfavorable external events (such as climate and environment) as well as measurement errors; the overall effect of these disturbances is unpredictable, but one expects the average effect of these factors to be negligible. Thus, define the probability spaces $(\mathbb{R}^+, \mathcal{F}, \mathbb{P}_{F|\mathbf{X}_1}(;\theta_1))$ and $(\mathbb{R}^+, \mathcal{F}, \mathbb{P}_{Q|\mathbf{X}_1}(;\theta_1))$, where $\mathcal{F}$ is the σ-field of (Borel) subsets of $\mathbb{R}^+$ and $\mathbb{P}_{F|\mathbf{X}_1}(.;\theta_1)$ and $\mathbb{P}_{Q|\mathbf{X}_1}(.;\theta_1)$ are associated conditional (on values of $\mathbf{X}_{1i}$) probability measures on $\mathcal{F} \times \Theta_1$ with conditional expected value $\rho(\mathbf{x}_{1i}; \beta_1)$, where Θ_1 is a certain parameter space augmenting $\mathcal{B}_1$ by including, for example, shape or dispersion parameters. Then, $\rho(.)$ is further transformed by a mapping $F_i : \mathcal{F} \mapsto \mathbb{R}^+$ (in the continuous case) or $Q_i : \mathcal{F} \mapsto \mathbb{Z}^+$ (in the discrete case), such that, for any a, $F_i^{-1}([a, \infty); \beta_1)$, $Q_i^{-1}(a; \beta_1)$ are both in $\mathcal{F}$ -and where our notation makes explicitly the dependence on β_1 via $\rho(.)$. Thus, the production functions $F_i(\mathbf{x}_{1i}; \beta_1)$ -in the continuous case- and $Q_i(\mathbf{x}_{1i}; \beta_1)$ -in the discrete case- are random variables with expected value (conditional on $\mathbf{x}_{1i}$) determined by $\rho(\mathbf{x}_{1i}; \beta_1)$.

[2]For notational simplicity, we use the notation $\rho(.)$ for the conditional expected value of both the continuous and discrete cases. However, the mapping could be well different in each case. A similar simplification is done later on, when dealing with the conditional expected value of the inefficiency term

Inefficiency is the second component explaining observed output levels. Equations (3) and (6) imply that $D(.)$ and $D^*(.)$ measure the maximal feasible contraction in the amounts of output for a *given* level of input. Let $\mathbf{X}_{2i}$ be the k_2 vector containing the subset of elements of $\mathbf{X}_i$ explaining the inefficiency level at unit i, and let $\beta_2 \in \mathcal{B}_2$ be the k_2 unobservable vector summarizing the contribution to inefficiency of each element of $\mathbf{X}_{2i}$. Conditional on $\mathbf{X}_{2i}$, the mean level of inefficiency at unit i would be summarized by a deterministic mapping, say $\eta : \mathbb{R}^{k_2} \times \mathcal{B}_2 \mapsto \mathcal{I} \subseteq \mathbb{R}^+$ transforming inputs in $\mathbf{X}_{2i}$, into inefficiency. As it happens with the production function, efficiency may itself be subject to random variation. In the discrete case (where $\mathcal{I}$ coincides with $\mathbb{R}^+$), randomness of inefficiency is captured by the mapping $D_i^* : \mathcal{F} \mapsto \mathbb{Z}^+$ (such that,for any a, $D_i^{*-1}(a; \beta_2) \in \mathcal{F}$- in the probability space $(\mathbb{R}^+, \mathcal{F}, \mathbb{P}_{D^*|\mathbf{X}_2}(.; \theta_2))$, where $\mathbb{P}_{D^*|\mathbf{X}_2}(.; \theta_2)$ has conditional expected value given by $\eta(.)$. In the continuous case, a similar setting (but now watching that $\mathcal{I} = [1, \infty)$) allows us to introduce the random variable $D_i(\mathbf{x}_{2i}; \beta_2)$ defined in the corresponding probability space with conditional probability $\mathbb{P}_{D|\mathbf{X}_2}(.; \theta_2)$. As before, $\theta_2 \in \Theta_2$, where the parameter space is such that $\mathcal{B}_2 \subseteq \Theta_2$.

Finally, the stochastic version of equations (4) and (7) are given by

$$Y_i = F_i(\mathbf{x}_{1i}; \beta_1) D_i(\mathbf{x}_{2i}; \beta_2) \begin{cases} F_i(\mathbf{x}_{1i}; \beta_1) \sim \mathbb{P}_{F|\mathbf{X}_1}(f_i|\mathbf{x}_{1i}; \theta_1) \\ D_i(\mathbf{x}_{2i}; \beta_2) \sim \mathbb{P}_{D|\mathbf{X}_2}(d_i|\mathbf{x}_{2i}; \theta_2) \end{cases} \tag{8}$$

$$Y_i = Q_i(\mathbf{x}_{1i}; \beta_1) + D_i^*(\mathbf{x}_{2i}; \beta_2) \begin{cases} Q_i(\mathbf{x}_{1i}; \beta_1) \sim \mathbb{P}_{Q|\mathbf{X}_1}(q_i|\mathbf{x}_{1i}; \theta_1) \\ D_i^*(\mathbf{x}_{2i}; \beta_2) \sim \mathbb{P}_{D^*|\mathbf{X}_2}(d_i^*|\mathbf{x}_{2i}; \theta_2) \end{cases} \tag{9}$$

An statistical model for each case may be now built by careful selection of probability laws for each of the random variables, and Maximum Likelihood may be employed for estimating the parameters of the model. However, attention need to be paid only to the latter model: equation (8) is similar to Stevenson's ([Stevenson, 1980]) version of the Stochastic Frontier Model firstly treated by Aigner, Lovell and Schmidt ([Aigner et al., 1977]) and Meeusen and van den Broeck ([Meeusen & van den Broeck, 1977]). The relationship between model (8) and these authors' models is explained, for example, in Kumhakar and Lovell ([Kumbhakar & Knox Lovell, 2003]). A logarithmic transformation Y_i leads to $\log Y_i = \log(F_i(\mathbf{x}_{1i}; \beta_1)) + \log(D_i(\mathbf{x}_{2i}; \beta_2))$ and probability functions may be associated to the transformations of $F(.)$ and $D(.)$. Often, one encounters that researchers' favorite distribution for $\log(F(.))$ is a Normal distribution with center determined by the levels of inputs; several choices have been made for the distribution of the log-inefficiency: the original paper of Aigner, Lovell and Schmidt considered $\log(D_i(\mathbf{x}_{2i};\beta_1))$ to be distributed as a Half Normal Distribution while Greene ([Greene, 1990]) considered a Gamma Distribution instead; exponential and truncated normal distributions have also been considered by others. For comments on these choices and their performance see Kumbhakar and Lovell. In general, one expect White's remarks on model selection (White, [White, 1982]) to apply.

3.1 *Discrete Outputs*

The case of outputs measured in discrete amounts had not been treated up to now, and therefore it is the second main contribution of this article. From equation (), the output level Y_i is ruled by the convolution of the two random variables $Q(.)$ and $D^*(.)$. Therefore, for $\Theta = \Theta_1 \bigcup \Theta_2$, the probability mass function of Y_i given $\mathbf{x}_i = \mathbf{x}_{1i} \bigcup \mathbf{x}_{2i}$ is

$$\begin{aligned}
\mathbb{P}_{Y_i|\mathbf{X}_i}(y_i|\mathbf{x}_i;\theta) &= \sum_{d_i^*=0}^{y_i} \mathbb{P}_{Q_i|\mathbf{X}_i}(y_i - u_i|\mathbf{x}_i;\theta) * \mathbb{P}_{D_i^*|\mathbf{X}_i}(d_i^*|\mathbf{x}_i;\theta) \\
&= \sum_{q_i=0}^{y_i} \mathbb{P}_{Q_i\mathbf{X}_i}(q_i|\mathbf{x}_{1i};\theta_1) * \mathbb{P}_{D_i^*|\mathbf{X}_i}(y_i - q_i|\mathbf{x}_{2i};\theta_2) \qquad (10)
\end{aligned}$$

provided that $Q_i(.)$ and $D_i^*(.)$ are independent random variables[3]. The parameters θ_1 and θ_2 of the model may be estimated by maximizing the likelihood function

$$\mathcal{L}(\beta,\gamma|\mathbf{Y},\mathbf{X}) = \prod_{i=1}^{n} \sum_{q_i=0}^{y_i} \mathbb{P}_{Q_i|\mathbf{X}_i}(q_i|\mathbf{x}_{1i};\theta_1) * \mathbb{P}_{D_i^*|\mathbf{X}_i}(y_i - q_i|\mathbf{x}_{2i};\theta_2) \qquad (11)$$

which only requires explicit definitions of the probability mass functions of $Q(.)$ and $D^*(.)$. The resulting maximum likelihood estimates, $\hat{\theta}_1$ and $\hat{\theta}_2$ may be used to obtain predicted values for $Q_i(.)$ and $d_i^*(.)$, namely $\hat{Q}_i(.) = \rho(\mathbf{x}_{1i};\hat{\beta}_1)$ and $\hat{D}_i^*(.) = \eta(\mathbf{x}_{2i};\hat{\beta}_2)$. As it happens with the typical stochastic frontier model, however, in order to obtain estimates of the inefficiency levels an estimator may be built from the posterior expectation of D_i^* as suggested by Jondrow, Lovell, Materov and Schmidt ([Jondrow et al., 1982]). Following

[3]This assumption is clearly restrictive, although as [Kumbhakar & Knox Lovell, 2003] point out in the monograph, some authors have found it *innocuous*. Probably because of this, the assumption underlies most of the literature on production frontiers; only recently Smith ([Smith, 2007]) has proposed the use of copula methods in order to relax the restriction. Although we conjecture that a similar approach could be adopted for our model, we leave the development of such device for later research.

these authors we have:

$$\begin{aligned}
\hat{D}^*(y_i, \mathbf{x}_i; \theta) &= E\left(D_i^* | y_i, \mathbf{x}_i\ ; \theta\right) \\
&= \sum_{d_i^*=0}^{\infty} d_i^* \mathbb{P}_{D_i^*|Y_i,\mathbf{X}_i}\left(d_i^* | y_i, \mathbf{x}_i; \theta\right) \\
&= \sum_{d_i^*=0}^{y_i} d_i^* \left\{ \frac{\mathbb{P}_{Y_i,D_i^*|\mathbf{X}_i}\left(\{Y_i = y_i\} \cap \{D_i^* = d_i^*\} | \mathbf{x}_i; \theta\right)}{\mathbb{P}_{Y_i|\mathbf{X}_i}\left(y_i | \mathbf{x}_i; \theta\right)} \right\} \qquad (12)
\end{aligned}$$

Clearly, $Y_i = y_i \Leftrightarrow Q_i = y_i - D_i^*$, so that

$$\begin{aligned}
E\left(D_i^* | y_i, \mathbf{x}_i; \theta\right) &= \sum_{d_i^*=0}^{y_i} d_i^* \left\{ \frac{\mathbb{P}_{Y_i,D_i^*|\mathbf{X}_i}\left(\{Q_i = y_i - d_i^*\} \cap \{D_i^* = d_i^*\} | \mathbf{x}_i; \theta\right)}{\mathbb{P}_{Y_i|\mathbf{X}_i}\left(y_i | \mathbf{x}_i; \theta\right)} \right\} \\
&= \sum_{d_i^*=0}^{y_i} d_i^* \left\{ \frac{\mathbb{P}_{Q_i|\mathbf{X}_i}\left(y_i - d_i^* | \mathbf{x}_i; \theta_1\right) \mathbb{P}_{D_i^*|\mathbf{X}_i}\left(d_i^* | \mathbf{x}_i; \theta_2\right)}{\mathbb{P}_{Y_i|\mathbf{X}_i}\left(y_i | \mathbf{x}_i; \theta\right)} \right\} \\
&= \sum_{d_i^*=0}^{y_i} d_i^* w_i \qquad (13)
\end{aligned}$$

where the weights w_i will depend on the particular choice of marginal probability mass functions.

3.2 Mixed Poisson Models and the Delaporte Distribution

Equation (11) defines a whole family of models for the *discrete frontier*. Different choices of distributions for $Q(.)$ and $D^*(.)$ will lead to different models with their own idiosyncrasy, which should be studied separately. However, certain choices of distributions seem to arise naturally. This is the case of Mixed Poisson Models.

A random variable, Z, has a Mixed Poisson distribution with mixing parameter $\alpha \in \mathcal{A}$, if $f(z) = \int_{\mathcal{A}} \mathbb{P}_{Z|\alpha}(z|\alpha) f(\alpha) d\alpha$ while $\mathbb{P}_{Z|\alpha}(z|\alpha) \sim \text{Poisson}(\rho\alpha)$, where $f(\alpha)$ is the density function of the mixing parameter and ρ is predetermined. Unlike the baseline Poisson distribution, Mixed Poisson models adapt over(under)dispersion, and this frailty may then be understood as a form of unobserved heterogeneity. In the present context, the advantage of Mixed Poisson models will be precisely the ability to identify this heterogeneity as a source of variation distinct from inefficiency. This is a rather appealing feature of the models and it is not shared by frequently encountered stochastic frontier models for continuous data.

The latter models are such that estimates of inefficiency are confounded by unobserved heterogeneity. This is specially true for those models based on straightforward application of fixed or random effect panel data models. The exception to the rule is the true fixed and random effects models by Greene ([Greene, 2005]).

To proceed with the specification of the model, the mixed distribution may be allocated to either the frontier or the inefficiency terms, depending on the conjectures drawn by the researcher regarding where the heterogeneity dwells. Then, the remaining term may be simply assumed to follow a Poisson distribution, since heterogeneity has already been taken into account. Interestingly, the convolution resulting under these premises is also Mixed Poisson distributed: the mixing distribution is a shifted (by the average inefficiency) version of the original mixing distribution (see Karlis and Xekalaki [Karlis & Xekalaki, 2005]). Under this specification, the researcher may then obtain parameter estimates for the structural parts of the average frontier, the efficiency term and the heterogeneity, and studies of the moments of the model may be done within the general framework of Mixed Poisson distributions.

Among the family of mixed Poisson models, the most elemental choice is the so called Delaporte distribution (see [Delaporte, 1962], [Ruohonen, 1989] and Willmot and Sundt [Willmot & Sundt, 1989]). This model result from convoluting a Negative Binomial distribution with a Poisson distribution[4]. It is well known that the Negative Binomial distribution is, in fact, a Mixed Poisson model with gamma-distributed mixing parameter $\alpha \sim \text{Gamma}(\delta, \gamma)$ (see [Greene, 2004] or [Hausman et al., 1984]). For identification purposes,it is customary to assumed that $\alpha \sim \text{Gamma}(\delta, \delta)$ which implies that $E(\alpha) = 1$ and $var(\alpha) = \delta^{-1}$. This, in turn, may be interpreted in a regression vein: heterogeneity acts as a regression error with null expected value and vanishing effect as its variance (determined by δ) vanishes.

In order to facilitate ML estimation, we shall redefine $\delta = \exp(\xi)$, so as to satisfy the restriction $\delta > 0$. In our model, the Delaporte distribution arises in two different ways. Firstly, one may assume that overdispersion is caused by the frontier term. Then, it follows

[4]Alternatively, the model may be obtained as a Mixed Poisson model with mixing parameter following a shifted gamma distribution (see Ruohonen [Ruohonen, 1989])

that (see for example, [Greene, 2004] or [Hausman et al., 1984])

$$\mathbb{P}_{Q_i|\mathbf{X}_i,\alpha_i}\left(q_i|\mathbf{x}_i,\alpha_i\right) = \begin{cases} \frac{e^{-\rho_i\alpha}(\rho_i\alpha)^{q_i}}{q_i!} & \text{for} \quad q = 0,1,\ldots \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$\begin{aligned} \mathbb{P}_{Q_i|\mathbf{X}_i}\left(q_i|\mathbf{x}_i\right) &= \int_0^\infty \mathbb{P}_{Q_i}\left(q_i|\mathbf{x}_i,\alpha_i\right) f\left(\alpha_i\right) d\alpha_i \qquad (15) \\ &= \frac{\Gamma\left(q_i+\delta\right)}{\Gamma\left(\delta\right) q_i!}\left(\frac{\delta}{\rho_i+\delta}\right)^{\delta}\left(\frac{\rho_i}{\rho_i+\delta}\right)^{q_i} \sim \mathcal{NB}\left(p_i,\delta\right) \end{aligned}$$

where $\mathcal{NB}\left(p_i,\delta\right)$ is the negative binomial density with parameter $p_i = \frac{\delta}{\rho_i+\delta}$, $\Gamma\left(.\right)$ is the gamma function and $\rho_i = \exp(\mathbf{x}_{1i};\beta_1)$. Then, given that the frontier $Q(.)$ has been assumed to be responsible for the heterogeneity in the data, we let D^* have a Poisson distribution (conditional on $\mathbf{x}_{2i}$) with mean parameter given by $\eta_i = \exp(\mathbf{x}_{2i}'\beta_2)$. The resulting Delaporte model for (10) is

$$\mathbb{P}_{Y_i|\mathbf{X}_i}\left(y_i|\mathbf{x}_i\,;\theta\right) = \sum_{q_i=0}^{y_i}\frac{\Gamma\left(q_i+\delta\right)}{\Gamma\left(\delta\right) q_i!}\left(\frac{\delta}{\rho_i+\delta}\right)^{\delta}\left(\frac{\rho_i}{\rho_i+\delta}\right)^{q_i}\frac{e^{-\eta_i}\eta_i^{(y_i-q_i)}}{(y_i-q_i)!} \tag{16}$$

This will be referred to as model DEL-1. Under the assumption of independence of Q_i and D_i^*, it follows that $E\left(Y_i|\mathbf{X}_i\right) = \rho_i + \eta_i$ and $var\left(Y_i|\mathbf{X}_i\right) = \rho_i\left(\frac{\rho_i}{\delta}+1\right) + \eta_i$ -both the sum of the conditional expected values and variances of the individual random variables. For this model, the efficiency measure in (13) has weights given by

$$w_i = \frac{\frac{\Gamma\left(y_i-d_i^*+\delta\right)}{\Gamma(\delta)\left(y_i-d_i^*\right)!}\left(\frac{\delta}{\rho_i+\delta}\right)^{\delta}\left(\frac{\rho_i}{\rho_i+\delta}\right)^{y_i-d_i^*}\frac{e^{-\eta_i}\eta_i^{\left(d_i^*\right)}}{d_i^*!}}{\sum_{d_i^*=0}^{y_i}\frac{\Gamma\left(y_i-d_i^*+\delta\right)}{\Gamma(\delta)\left(y_i-d_i^*\right)!}\left(\frac{\delta}{\rho_i+\delta}\right)^{\delta}\left(\frac{\rho_i}{\rho_i+\delta}\right)^{y_i-d_i^*}\frac{e^{-\eta_i}\eta_i^{\left(d_i^*\right)}}{d_i^*!}} \tag{17}$$

The coefficient associated to the covariate $X_{ij} \in \mathbf{X}_i = \mathbf{X}_{1i}\bigcup\mathbf{X}_{2i}$, say β_j satisfies that

$$\frac{\partial E(Y_i|\mathbf{X}_i)}{\partial X_{ij}} = \beta_j(e^{\mathbf{x}_{1i}'\beta_1} + e^{\mathbf{x}_{2i}'\beta_2}) \Rightarrow \beta_j = \frac{\partial E(Y_i|\mathbf{X}_i)}{\partial X_{ij}}\frac{1}{E(Y_i|\mathbf{X}_i)}$$

which can be interpreted as an elasticity (capturing the percentage variation in the expected value of Y when the j^{th} input changes by ∂X_j percent). If the associated variable were categorical, then, is is more appropriate to consider the effect of varying the regressor of interest by one unit. Suppose, for example, $E(Y_i|\mathbf{X}_i) = e^{\mathbf{X}_{1i}'\beta_1} + e^{\mathbf{X}_{2i}'\beta_2+Z_i\xi}$, where Z_i is a categorical variable and $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ have no elements in common. Then,

$$\begin{aligned}
\Delta E(Y_i|\mathbf{X}_i) &= e^{\mathbf{X}_{1i}'\beta_1} + e^{\mathbf{X}_{2i}'\beta_2+(Z_i+1)\xi} - (e^{\mathbf{X}_{1i}'\beta_1} + e^{\mathbf{X}_{2i}'\beta_2+Z_i\xi}) && (18)\\
&= e^{\mathbf{X}_{2i}'\beta_2} e^{Z_i\xi}(e^{\xi}-1) = E(D_i^*|\mathbf{X}_i)(e^{\xi}-1) && (19)\\
\xi &= \log\left(1+\frac{\Delta E(Y_i|\mathbf{X}_i)}{E(D_i^*|\mathbf{X}_i)}\right) && (20)\\
&= \frac{\Delta E(Y_i|\mathbf{X}_i)}{E(D_i^*|\mathbf{X}_i)} + O\left\{\left(\frac{\Delta E(Y_i|\mathbf{X}_i)}{E(D_i^*|\mathbf{X}_i)}\right)^2\right\} && (21)
\end{aligned}$$

via a Taylor series expansion. Thus, ξ may be approximately interpreted as the expected ratio of variation in Y given the variation in Z_i.

In the model DEL-1 all the unobserved heterogeneity is concentrated in the frontier. Alternatively, one may assume that overdispersion would be a natural feature of inefficiency $D^*(.)$, while the differences in the frontier would be driven by observed information an pure statistic noise. This alternative model, which will be referred to as DEL-2 would be given by

$$\mathbb{P}_{Y_i|\mathbf{X}_i}(y_i|\mathbf{x}_i\,;\theta) = \sum_{q_i=0}^{y_i} \frac{\Gamma(y_i-q_i+\delta)}{\Gamma(\delta)(y_i-q_i)!}\left(\frac{\delta}{\eta_i+\delta}\right)^{\delta}\left(\frac{\eta_i}{\eta_i+\delta}\right)^{y_i-q_i}\frac{e^{-\rho_i}\rho_i^{q_i}}{q_i!} \tag{22}$$

For this model the weights of the efficiency measure (13) would be an obvious modification of those in (17).

3.3 Near Identification

Implementation of the above models would follow from maximization of the corresponding likelihood functions. However, from a empirical perspective, there is a potential problem with the identification of the parameters of the Delaporte model. The Negative Binomial part of the model may collapse to a Poisson law whenever $\delta \to \infty$, so that $var(\alpha) \to 0$. Therefore, in the limit one ends up with the convolution of two independent Poisson distributions. Thus if $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ contain only intercept terms, we may not find a unique estimate of the individual coefficients of each intercept. Similarly, if $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ contain a single common regressor, there is not unique solution for the score equation of the maximum likelihood function: if $Y \sim \text{Poisson}(a = e^{x\beta} + e^{x\gamma})$, then $E(Y|x) = a = e^{x\beta} + e^{x\gamma}$ does not admit a unique solution, so that similar distributions may be generated by using different parameter values.

In general, the problem we face is one of *Near Identification*, similar to those affecting the traditional ALS-type model (see [Bandyopadhyay & Das, 2008]). Bandyopadhyay and Das define near-identification as a local loss of curvature in the likelihood function, so that parameter values relatively far from the true parameter value may generate values of the likelihood function arbitrarily close to the optimum. As discussed in that article, typical cross-sectional SFM with uncorrelated error components suffer from near identifiability problem. In their opinion this accounts, then, for the lack of precision reported by [Ritter & Simar, 1997] of the many variations of the ALS model. Thus, typical algorithms for optimization will provide less precise estimates.

The Delaporte models are also subject to this loss of curvature, which becomes apparent as δ increases in value. Thus in model DEL-1 we observe that as $\delta \to \infty$,

$$\begin{aligned}
\frac{\partial \mathbb{P}_{Y_i|\mathbf{X}_i}(y_i|\mathbf{x}_i;\theta)}{\partial \delta} &= \sum_{q_i=0}^{y_i} \mathcal{P}(\eta_i)\mathcal{NB}(p_i,\delta) \\
&\times \left(\frac{-q_i}{\rho_i+\delta} + \ln\left(\frac{\delta}{\rho_i+\delta}\right)\frac{\rho_i}{(\rho_i+\delta)^2} + \frac{\dot{\Gamma}(q_i+\delta)}{\Gamma(q_i+\delta)} - \frac{\dot{\Gamma}(\delta)}{\Gamma(\delta)} \right) \\
&\to 0
\end{aligned} \tag{23}$$

where $\dot{\Gamma}(.)$ is the derivative of the gamma function. The result is obvious once we consider that for any z,

$$-\frac{\dot{\Gamma}(z)}{\Gamma(z)} = z^{-1} + c + \sum_{n=1}^{\infty}\left[\frac{1}{n+z} - \frac{1}{n}\right]$$

where c is the Euler-Mascheroni constant[5]. Very large values of δ will cause the log-likelihood to be very flat in the direction of its feasible set, so that traditional algorithms for optimization will struggle to find the optimum. It is convenient to remark that because the problem is due to a fall in the variance of the mixing parameter in the Mixed Poisson part of the convolution. It is thus evident that the situation is not a peculiarity of the Delaporte model, and it will arise whenever we convolute a Poisson and a Mixed Poisson random variables.

[5]See Wolfram's `http://mathworld.wolfram.com/GammaFunction.html` and references therein

3.4 Monte Carlo

The magnitude of the identification problem, as well as the small sample merits of the Delaporte distribution, may be evaluated via a simple Monte Carlo experiment. Data was generated from the two Delaporte specifications discussed so far. The structural part of the models were $\rho_i = \exp(\beta_{1,0} + \beta_{1,1} X_{1i})$ and $\eta_i = \exp(\beta_{2,0} + \beta_{2,1} X_{2i})$, where $X_{1i} \sim U(0, 2)$ and $X_{2i} = 0.8Z + 0.9 * X_{1i}$ (which induced a somehow strong multicollinearity) and Z is a standard normal random variable. The values of the parameters were set to $\beta_{1,0} = \beta_{2,0} = 1$ and $\beta_{1,1} = \beta_{2,1} = 0.5$. The parameter $\delta = \exp(\xi)$ (which measured the degree of overdispersion) was set at three different values $\exp(\xi) = 0.3678$, 1, and e. The largest the value of δ, the lower the level of overdispersion caused by the negative binomial part of the model, the lower the variance of the mixing parameter and the likelier is the problem of near identification to arise. Samples of varying size were drawn from the generating processes DEL-1 and DEL-2, and each sample was estimated by maximizing the likelihood function associated to models DEL-1 and DEL-2. The experiment was repeated 200 times, and in each occasion, the Mean Square Error (MSE) of the estimated parameters was retained. At the end, the average MSE was calculated, and the results for $N = 1000$ are collected in Table 1.

Table 1 about here

The first and fourth columns of the table capture the MSE when the estimated model replicated the underlying DGP, while columns two and three collect the results under misspecification. Under correct specification of the model the reported MSE is small, as expected given the general theory of maximum likelihood. Reported MSE oscillate about 5% of the true value of the parameter, but larger values of δ do increase de MSE of the parameters in the structural parts of the model. The increase is not as large as to substantially compromise the estimated values; in fact, simulations not reported here suggest that a value of ξ above 4 is required to observe compromising losses of accuracy. Under correct specification, we observe that variance and square bias share equal proportions of the MSE, so that, loss of precision with δ is due to increases in bias and variance in the same proportion. Among the two models, DEL-1 seems to be more sensitive to the variation of δ than DEL-2.

Under model misspecification (columns 2 and 3), simulations report larger MSE whose magnitudes are driven by the square bias -as expected. However, it seems that DEL-2 is more sensitive to model misspecification than DEL-1. This may be due to the fact that the base line model for the efficiency frontier is a Poisson model which imposes restrictive moment conditions reducing the admitted amount of dispersion in the sample.

These results would suggest that, in practice, the estimated value of δ must be observed as a measure of the accuracy with which the parameters of the model are estimated; for equally low/moderate ξ researchers may rely on the robustness of DEL-1 for their inferences.

A single draw of each of the above models was used in order to get an indication of the ability of the efficiency measure in equation (13) to retrieve the actual density of the inefficiency. The values of $D^*(.)$ generated by the computer and those values generated by the DEL-1 and DEL-2 versions of equation (13) were retained, and the Nadaraya-Watson estimators of these values where plotted against each other. Each of the panels in Figures 1 and 2 collect one such comparison for values of $\xi = -1, 1,$ and 3 -with $\beta_{.,.}$ as above.

[Figure 1 about here]

[Figure 2 about here]

The estimator seems to provide sufficiently accurate estimates of the underlying distribution (under no misspecification), even for situations when a notable loss of identification is suspected. Values provided by the estimator seem to be slightly more concentrated about the mean, so that the estimator tends to be slightly conservative and this seems to become more so the closer the DGP is to the convolution of two Poisson distributions (large ξ). Nonetheless, even for $\xi = 3 \rightarrow \delta = exp(3) \cong 20$, and $var(\alpha) \cong 0.04$ estimates seem sufficiently reliable.

Finally it is remarked that the method is rather robust to misspecification in the distribution of the mixing parameter of the Mixed Poisson part of the distribution of Y. Simulations not reported here show that, provided that the distribution of α has unit expected value, the ML method returns accurate estimates of the parameters of the model as well as estimates of the inefficiency. This is particularly true for model DEL-2. Experiments with $\chi^2(1)$, $Exponential(\ln(2))$ and $Weibull(1,1)$ all result in successful estimation of the conditional mean and distribution of the efficiency. The exception to this was the case when α followed a $Uniform(0,2)$ distribution, which resulted in a somehow imprecise estimates of the distribution of the efficiency.

4 Application: Infant Mortality in England (2001)

In this section, models DEL-1 and DEL-2 are applied to a data set with the aim of exploring the drivers of infant mortality when aggregated at regional level. The geographical units of interest are 353 Local Authorities in England and the data refers to the year 2001.

Infant deaths are casualties occurring when a baby is under one year old. It is conjectured that the expected number of infant deaths in a local authority is largely determined by population size. However, beyond the role of population size there are a number of factors which are likely to cause observed counts to exceed the expected frontier levels. In particular these factors may be summarized in three categories: environmental, economic and educational.

Environmental factors have been associated to the overall health of the population (see, for example, Tulchinsky and Varavikova [Tulchinsky & Varavikova, 2000], chapter 9). In particular high emissions of pollutants to the atmosphere can induce respiratory diseases such as bronchitis, pneumonia, allergies or asthma, as well as interfere with hemoglobin oxygen carrying capacity, which can seriously affect fetal development. This is the case of Nitrogen Oxides (NO_2, Nitrogen Dioxide, and NO, Nitric Oxide), which are by-products of fuel consumption and the production of electricity. It is estimated that 50% of emisions of these gases are due to road traffic, while another 20% is due to the production and consumption of electricity.

Education and economic deprivation are also likely drivers of inefficiency in the production of infant deaths. The latter may explain the population's ability to access healthy food and life styles which repercute in the baby's development while in uterus and later life. Education levels, in turn, are likely to capture the population's exposure to elementary baby-caring measures, as well as its ability for risk management. For example, it is sensible to conjecture that a well informed population might eliminate more successfully risks leading to Sudden Infant Deaths (SID) (often avoidable casualties such as asfixia during sleep or death due to hyperthermia).

A final factor that may be relevant in a analysis of infant mortality is the levels of pressure over medical and public services in the region, measurable through population density. In principle, one may argue that where population density is high, medical resources are under higher pressure, slowing the response of public services all things being equal. Slow response, in turn, may lead to failures to reduce avoidable deaths, with the consequent increase in reported counts of infant deaths.

To evaluate these conjectures, a data set was compiled with information downloaded from the website of the UK Office of National Statistics (ONS). Apart from the counts of infant deaths in the year 2001, the data set was augmented by including the following variables. Population levels in 2001, p_i was available, as well as the area (in square kilometers) for each local authority was collected; from these, we calculated population density, d_i as the ratio of the two. To capture environmental pollution, the ONS' Local Area Average of NO_x emissions intensity score, denoted NO_i^s was considered. This score corresponds to an 8 step scale, where each step represents an interval of emissions of NO_x in tonnes per square kilometer. Higher scores are associated to high levels of emissions. Economic deprivation was measured via the proportion of the area's active population receiving unemployment benefits, u_i, and education levels were proxied by the proportion of the population without academic qualifications, e_i.

4.1 Models and Results

The deterministic part of the Delaporte models is given by the following pair of equations:

$$\rho(\mathbf{x}_{1i};\beta_1) = \exp\left(\beta_{10} + \beta_{11}p_i + \beta_{12}\mathbb{I}_{Ni} + \beta_{13}\mathbb{I}_{Mi} + \beta_{14}\mathbb{I}_{Si}\right) \quad (24)$$

$$\eta(\mathbf{x}_{2i};\beta_2) = \exp\left(\beta_{20} + \beta_{21}NO_i^s + \beta_{22}u_i + \beta_{23}e_i + \beta_{24}d_i\right) \quad (25)$$

where $\mathbb{I}$ is an area identifier (North/Mid/South of England as well as City of London, which has been excluded) intended to capture other unobserved area characteristics affecting the levels of the frontier. The estimates of the parameters are reported in Table 2.

[Table 2 about here]

To assess the quality of the estimates, attention is firstly focused on the estimated value of the parameter ξ; the estimated values returned by DEL-1 and DEL-2 are 2.33 and 0.35 respectively (the latter significant only at 10%). This implies that the magnitudes reported by DEL-2 are going to be more accurate than those reported by DEL-1 (provided no model specification in involved).

Model DEL-1 presents an estimated frontier with a higher slope, which may be seen by looking at the lower level of the intercept and faster growing effect of population size; thus the model is penalizing less populated local areas with higher inefficiency scores. This may be seen also in Figures 3 and 4 showing the estimated frontiers from the models.

[Figure 3 about here]

[Figure 4 about here]

In the plots it is possible to infer the different frontiers for each of the four regions considered (North, Middle, South and City of London), although the estimated regional effect is only significant in the case of the South, where the expected number of deaths seems to be lower than elsewhere in the country. It is visually apparent that DEL-2 envelopes the data from below in a tighter fashion than DEL-1 does. Therefore, DEL-2 seems to provide an estimated frontier which is closer to the idea of a frontier for an economic bad. All this considered, model DEL-2 seems to be a better specification for the analysis.

The coefficients of the determinants of inefficiency are collected in the second part of the table. The models seem to suggest that there is a structural level of inefficiency, captured by the negative, but significant, intercept. All the factors considered in the model contribute to increase the inefficiency, as expected (however, note that the coefficient of NO score in DEL-2 is not significant). The models, however, do not agree on the magnitude of the effect of each variable. DEL-1 seems to suggest that economic deprivation has a less dramatic effect on inefficiency than education levels (coefficients of 0.02 and 0.04 respectively). In the view of this model, and incrase of 1 point in the NO score would lead to an increase in expected inefficiency of about 5% -in accordance to the expression for the marginal product given in the previous section. Therefore, DEL-1 suggests that enviromental pollution is the most important determinant of infant deaths. The conclusions drawn from DEL-2 are similar: 1 extra point in the NO score would lead to a 6% increase in the expected inefficiency; the second most important factor would be unemployment followed by education levels. Finally both models coincide to outline the importance of population density: all the things being equal, more densely populated areas seem to be more inefficient than less densely populated areas.

[Figure 5 about here]

Finally, Figure 5 presents kernel density estimates of the estimated inefficiencies for each of the models. The estimated density of the inefficiencies provided by DEL-2 has a much longer tail than the density returned by DEL-1. This reflects the well known fact that

DEL-2 allows for overdispersed counts in the inefficiency term, while DEL-1 imposes the restriction that the conditional mean of the inefficiency equals the conditional variance, so that very extreme counts are unlikely.

Overall, results seem to suggest that environmental policies are likely to have a larger impact on infant mortality that those policies oriented to unemployment or education. It is unclear, however, which factor, among unemployment or education, should be a priority. Since DEL-2 is likely to fit the data better than DEL-1, one would think that unemployment policies facilitating access to income are more likely to contribute to a reduction of infant mortality than other initiatives oriented to reduce education inequality or population agglomerations.

5 Conclusion

This article has introduced a model of stochastic frontiers applicable to those situation when the dependent variable is an economic bad and measured in discrete (non-negative) amounts. The need for such a model has been justified from a theoretical perspective, by noting that distance functions on which typical stochastic frontier models are based do not suit when the dependent variable is a count. The problem is the multiplicative nature of these schemes which implies that the observed quantity of an output is the result of augmenting the levels predicted by the production frontier by a factor larger than one (in the case of economic bads). If output is measured in discrete amounts, then an additive scheme is needed instead, to be able to capture the discrete nature of both production frontier and inefficiency levels.

This discussion led us to the new econometric model. As in ALS, the model is also the convolution of two random variables (one associated to the frontier and another associated to the inefficiency). The difference is that the associated random variables are discrete valued and therefore demand discrete probability distributions. Like in ALS, the models here presented are estimable at least via maximum likelihood methods, and efficiency may be estimated as in Jondrow et al, via the conditional posterior expectation of the random variable capturing inefficiency. However, unlike well known stochastic frontier models, ours can separate unobserved heterogeneity and inefficiency, so that estimates of efficiency are not contaminated. THe models in this article also share a feature of other likelihood based stochastic frontiers: the problem of *near identification* as discussed in Bandyopadhyay and Das. When the variance of the parameter causing the heterogeneity is too small, the log-likelihood becomes too flat about the true value of this parameter, causing a loss of precision of the estimates. This problem was reported by Ritter and Simar, and this article has tried to characterize its incidence via Monte Carlo simulations. In the view of the evidence presented here, however, values of δ beyond 19 are required in order to rises concerns about the quality of the estimates. From a statistical perspective our model is likely to be most useful when the range of values taken by the output variable is restricted to a few dozens (perhaps two or three hundred values). Beyond that, continuous approximations are likely to be still useful. In this sense, our model is not different to standard count data models (see comments in Hausman, Hall and Griliches).

References

[Aigner et al., 1977] Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6, 21–37.

[Bandyopadhyay & Das, 2008] Bandyopadhyay, D. & Das, A. (2008). Identifiability of the stochastic frontier models. *Munich Personal RePEc Archive*, No. 8032.

[Blackorby et al., 1978] Blackorby, C., Primont, D., & Russell, R. (1978). *Duality, Separability and Functional Structure: Theory and Economic Applications*. Elsevier North-Holland.

[Delaporte, 1962] Delaporte, P. (1962). Sur l'efficacite des criteres de tarification de l'assurance contre les accidents d'automobiles. *ASTIN Bulletin*, 2, 84–95.

[Fare et al., 1994] Fare, R., Grosskopf, S., & Lovell, C. (1994). *Production Frontiers*. Cambridge University Press.

[Greene, 1990] Greene, W. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46, 141–163.

[Greene, 2004] Greene, W. (2004). *Econometric Analysis*. Prentice Hall.

[Greene, 2005] Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126, 269–303.

[Hausman et al., 1984] Hausman, J., Hall, B., & Griliches, Z. (1984). Econometric models for count data with an application to the patents rd relationship. *Econometrica*, 52, 909–938.

[Jondrow et al., 1982] Jondrow, J., Materov, I., Lovell, K., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19, 233–238.

[Karlis & Xekalaki, 2005] Karlis, D. & Xekalaki, E. (2005). Mixed poisson distributions. *International Statistical Review*, 73, 35–58.

[Kumbhakar & Knox Lovell, 2003] Kumbhakar, S. C. & Knox Lovell, C. (2003). *Stochastic Frontier Analysis*. Cambridge University Press.

[Meeusen & van den Broeck, 1977] Meeusen, W. & van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review*, 18, 435–444.

[Ritter & Simar, 1997] Ritter, C. & Simar, L. (1997). Pitfalls of norma-gamma stochastic frontier models. *Journal of Productivity Analysis*, 8, 167–182.

[Ruohonen, 1989] Ruohonen, M. (1989). On a model for the claim number process. *ASTIN Bulletin*, 18, 57–68.

[Shephard, 1970] Shephard, R. (1970). *Theory of Cost and Production Function*. Princeton University Press.

[Smith, 2007] Smith, M. (2007). Stochastic frontier models with dependent error components. *The University of Sydney*, MIMEO.

[Stevenson, 1980] Stevenson, R. (1980). Likelihood function for generalized stochastic frontier estimation. *Journal of Econometrics*, 13, 57–66.

[Tulchinsky & Varavikova, 2000] Tulchinsky, T. & Varavikova, E. (2000). *The New Public Health*. Academic Press.

[White, 1982] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 53, 1–16.

[Willmot & Sundt, 1989] Willmot, G. & Sundt, B. (1989). On posterior probabilities and moments in mixed poisson processes. *Scandinavian Actuarial Journal*, 14, 136–146.

Mean Square Error

DGP	DEL-1						DEL-2					
Model	DEL-1			DEL-2			DEL-1			DEL-2		
	M.S.E.	*% B^2*	*Var.*	*M.S.E.*	*% B^2*	*Var.*	*M.S.E.*	*% B^2*	*Var.*	*M.S.E.*	*% B^2*	*Var.*
β_{10}	0.0303	0.501	0.015	0.2485	0.9	0.0248	0.0387	0.56	0.0169	0.00915	0.507	0.0045121
β_{11}	0.0204	0.504	0.0101	0.0123	0.52	0.006	0.0424	0.75	0.0103	0.00418	0.506	0.00206
$\xi = -1$	0.0345	0.501	0.0172	0.6236	0.95	0.0288	0.0904	0.69	0.0279	0.0312	0.505	0.0154
β_{20}	0.00781	0.5	0.0039	0.2126	0.97	0.0061	0.0608	0.92	0.0047	0.0174	0.5	0.0087
β_{21}	0.0014	0.501	0.00069	0.0214	0.88	0.0024	0.0831	0.98	0.00098	0.0075	0.5	0.00377
β_{10}	0.04005	0.5	0.02	2.9416	0.85	0.439	0.0377	0.51	0.0181	0.0193	0.5	0.0964
β_{11}	0.01257	0.5	0.00628	0.3556	0.75	0.0888	0.0243	0.77	0.0054	0.0056	0.5	0.0028
$\xi = 0$	0.0801	0.501	0.04	1.044	0.96	0.035	0.0982	0.56	0.042	0.056	0.5	0.028
β_{20}	0.0342	0.5	0.0171	0.442	0.98	0.00841	0.0472	0.64	0.0166	0.0266	0.502	0.0132
β_{21}	0.003889	0.5	0.0019	0.028	0.96	0.00104	0.0386	0.93	0.002338	0.00432	0.502	0.002154
β_{10}	0.0796	0.5	0.039729	3.178	0.865	0.428	0.13136	0.5	0.0648	0.05931	0.502	0.029501
β_{11}	0.0114	0.5	0.00572	0.4907	0.807	0.0945	0.0222	0.57	0.0093	0.007709	0.501	0.00384
$\xi = 1$	0.157	0.5	0.0787	0.8995	0.969	0.027425	0.24709	0.5	0.123	0.11858	0.501	0.058102
β_{20}	0.0803	0.501	0.04	0.4246	0.976	0.00999	0.161	0.5	0.0808	0.0647	0.501	0.032316
β_{21}	0.00707	0.501	0.0035	0.029041	0.9689	0.0009	0.0186	0.56	0.00805	0.00618	0.501	0.0030813

Table 1: *Replications* $R = 200$*; Sample Size* $N = 1000$; $(\beta_{10} = 1, \beta_{11} = 0.5, \beta_{20} = 1, \beta_{21} = 0.5)$ and varying ξ

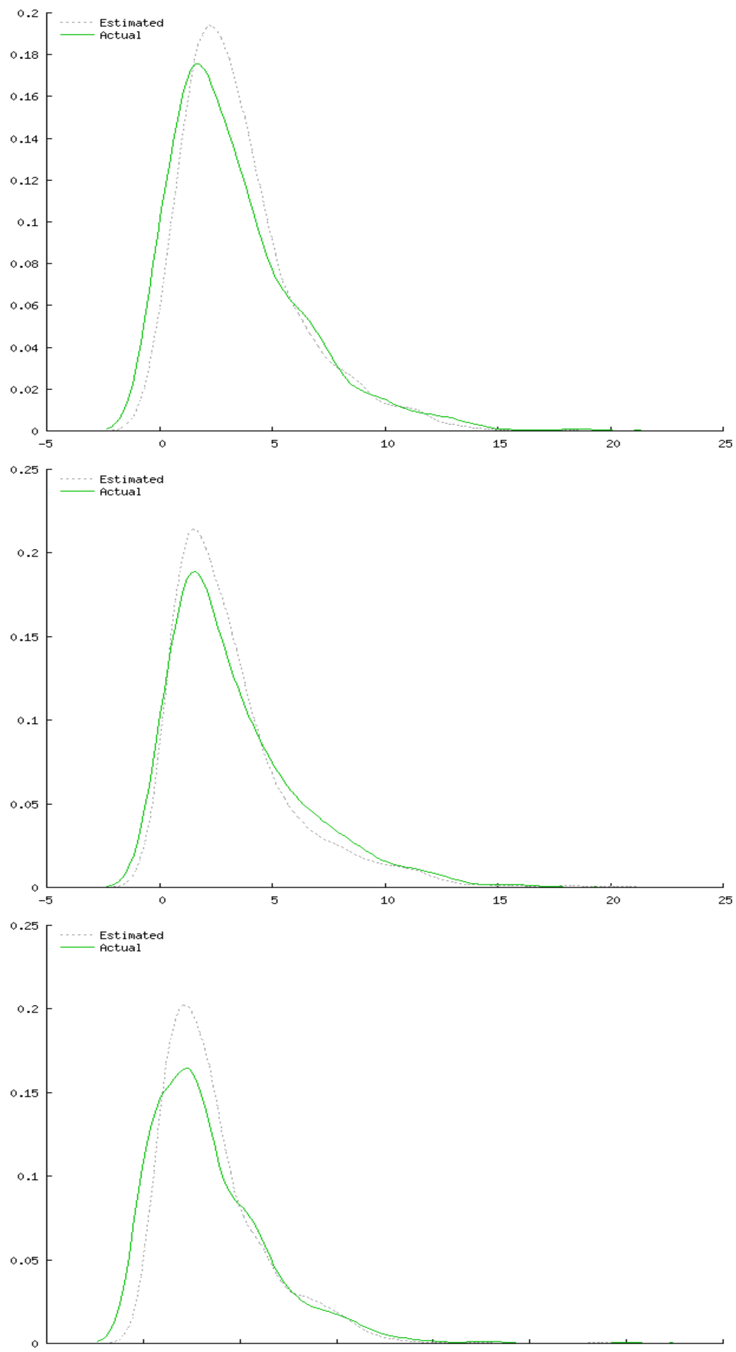

Figure 1: One-shot smoothed histograms of estimated and actual efficiency mass function for model DEL 1: $(\beta_{10} = 1, \beta_{11} = 0.5, \beta_{20} = 1, \beta_{21} = 0.5)$ and $\xi = -1$ (top figure),$\xi = 1$ (middle figure) and $\xi = 3$

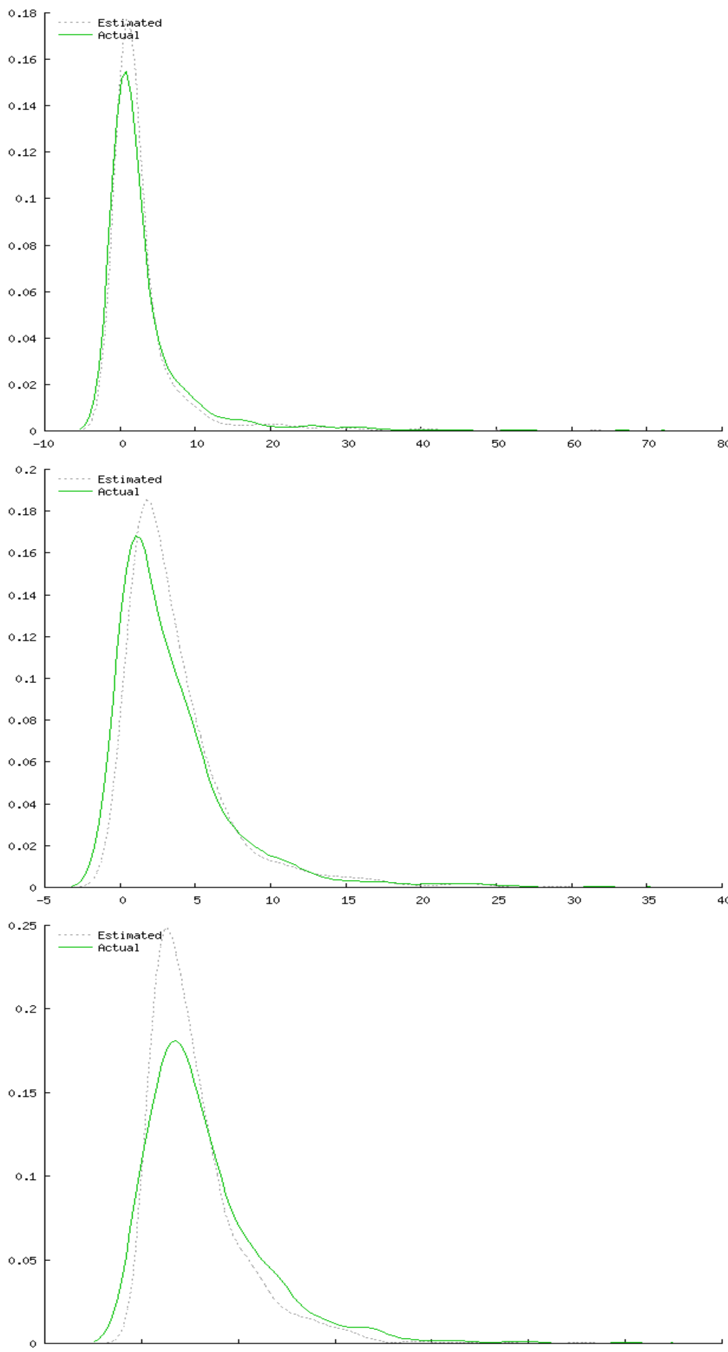

Figure 2: One-shot smoothed histograms of estimated and actual efficiency mass function for model DEL 2: $(\beta_{10} = 1, \beta_{11} = 0.5, \beta_{20} = 1, \beta_{21} = 0.5)$ and $\xi = -1$ (top figure),$\xi = 1$ (middle figure) and $\xi = 3$

	DEL-1		DEL-2	
	Coefficient	**S.E.**	**Coefficient**	**S.E.**
Frontier				
Constant	-16.134	(0.97161)	-14.146	(0.60275)
Population	1.6086	(0.077672)	1.4344	(0.047355)
North	0.007454	(0.090626)	-0.057702	(0.079731)
Central	-0.024168	(0.089301)	-0.016373	(0.077307)
South	-0.13333	(0.089301)	-0.16373	(0.079082)
ξ	2.3348	(0.15871)	0.35926	(0.21529)
Inefficiency				
Constant	-1.1501	(0.54148)	-1.2899	(0.52908)
Benefits	0.026299	(0.01488)	0.056498	(0.022244)
Non-qualified	0.042480	(0.010600)	0.040188	(0.012695)
Mean NO Score	0.052355	(0.028217)	0.064117	(0.05947)
Pop. Density	0.0222	(0.0027594)	0.026281	(0.00044242)

Table 2: Estimates of parameters of the model. The dependent variable is the count of infant deaths at Local Authority Level. $N = 353$

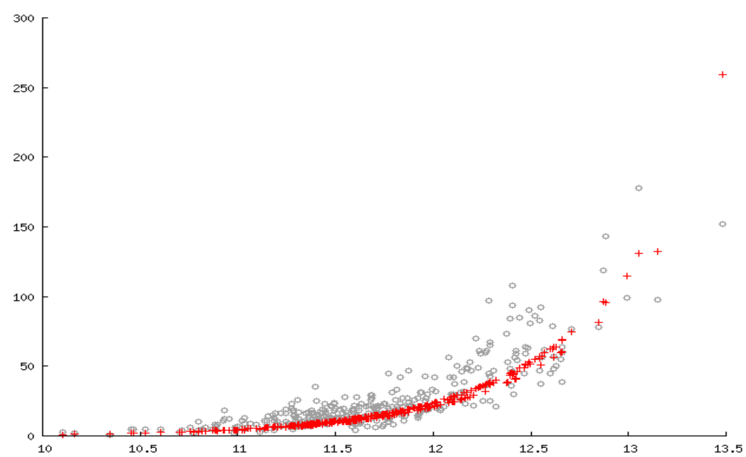

Figure 3: Estimated Frontier: Model DEL-1

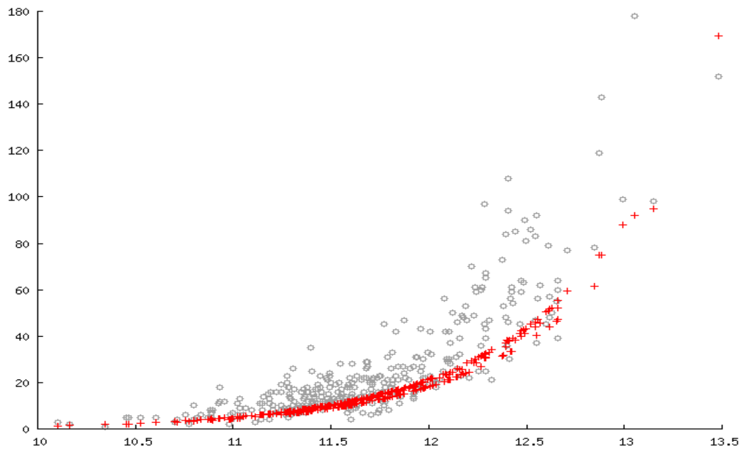

Figure 4: Estimated Frontier: Model DEL-2

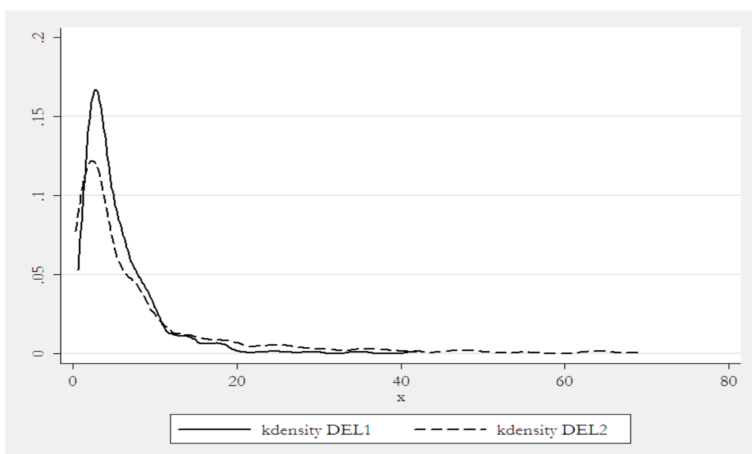

Figure 5: Estimated Densities for the Infant Mortality Model (DEL-1 solid line, DEL-2 dashed line)