MANCHESTER 1824
The University of Manchester

Economics
Discussion Paper Series
EDP-0725

Exploring a stochastic frontier model when the dependent variable is a count

Eduardo Fé-Rodríguez

November 2007

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

Exploring a Stochastic Frontier Model 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

November 19, 2007

Abstract

Stochastic Frontier Models have been successfully implemented in the estimation of short-run cost curves and the measurement of productive efficiency. Although developed and used mostly by economists, the model can be applied in the realm of other disciplines such as Epidemiology or Public Health by adequately re-interpreting the ideas of cost and productive efficiency. In these areas, cost is generally interpreted as a count of events, such as number of deaths or observed numbers of certain disease. Traditional SFM assume that the dependent variable under study has a continuous distribution (generally resulting from a convolution of continuous distributions). Therefore application of these models neglect the discrete nature of the variables of interest. In this article a new Maximum Likelihood model is proposed to estimate Stochastic Frontier Models with a discrete valued dependent variable. Measures of cross-sectional levels of efficiency are constructed through recurring to the conditional expected value of the inefficiency term given the information encompassed by the observed count.

Key Words: Stochastic Frontier, Discrete Random Variable, Discrete Convolution, Delaporte Distribution.

1 Introduction

In this article, we present a likelihood based Stochastic Frontier Model which can estimate *cost* functions whenever the dependent variable is a discrete valued random variable. The need for such a model arose from a study of the determinants of infant mortality in England. Our dependent variable was a count of the total number of newborn children who died prematurely (before one year of age) in England in 2001. The aim of the study was (1) to identify the drivers of absolute mortality and (2) to identify those areas of relatively high mortality. Accurate estimation of cost frontiers is a topic of importance for researchers and policymakers. In economics, a good behavioural approximation to the curve of production/cost possibilities provides information about the optimal output/cost attainable with a given technology. It also informs about the marginal product of each input in the process and the marginal rates of substitution between inputs. Likewise, in a health environment as the one described later in the paper, estimation of the *cost* frontier will assist us to (1) measure the minimal expected outcome of certain undesirable condition (2) to understand the marginal contribution or responsibility of each harmful agent in the health production process and (3) to approximate the marginal *rate of substitution* among these harmful agents. This latter point is important in the development of public health measures, since inferences about this rate of substitution can assist to discriminate among policies affecting different epidemiological vectors carrying or transmitting undesirable effects from the environment to a group of hosts or to society (as well as substituting between policies to deliver a more cost effective public health).

Although the study of frontiers is intrinsically united to the economics of production, the concept itself is an accomplished one, so that efficiency analysis can greatly contribute to a range of other seemingly unrelated disciplines such as epidemiology and public health. However, successful exportation of the idea of frontier often requires a reinterpretation of the concepts of efficiency and cost, the latter being frequently found in terms of count of events.

Following the literature on health production, we could characterize a (mortality) frontier quantifying the minimal expected number of infant deaths per year per area. The actual number of infant deaths registered at a geographic unit would depart from that minimal number due to two different sources of *variation*. Firstly, the area might be subject to unpredictable *shocks*, such as natural disasters or luck, which may cause unusually high (or lower than expected) numbers of deaths. In other words, the *mortality frontier* is stochastic across geographies. Secondly, a given area might be subject to idiosyncratic, albeit unobservable environmental effects which may lead to higher mortality than *optimal*; in the economic literature on production and cost frontiers, the latter shocks are referred to as technical efficiency.

This setting we have just described is similar to that encountered in economic applications of SFM. However, there are two major differences with respect to this traditional setting. Firstly, the ideas of cost function and technical productive efficiency needed to be accommodated to a new scenario: cost is now measured in terms different to monetary units, and efficiency is studied in terms of additional-to-expected lives losses. Secondly, number of deaths is a discrete valued random variable with certain mass function. Existing models for frontier analysis are built around continuously distributed random variables and, therefore, will fail to capture the discrete nature of the dependent variable, yielding inefficient and inconsistent estimations.

The model developed in this paper tries to solve the above problem. It is inspired on earlier work on applications of the Poisson Distribution by Hausman, Hall and Griliches ([7]) and Cameron and Trivedi ([2], [3]), as well as pioneer work on Stochastic Frontier Analysis by Aigner, Lovell and Schmidt ([1]), Meeusen and Broeck ([11]) and Greene ([5]). The model results from the discrete convolution of a Negative Binomial and a Poisson distributions. The former distribution appears as a Mixture Poisson random variable, designed to controlling two features: firstly, the stochastic variation of the frontier conditional upon a set of regressors and, secondly, unobserved heterogeneity in sample, so that individuals are allowed to differ randomly due to features of the data other than inefficiency, not fully accounted for by the regressors and unobservable to the researcher. The second component of the convolution, the Poisson Distribution, is used in other capture the inefficiency in the sample.

The convolution of a Poisson and Negative Binomial is known as a Delaporte Distribution. Ruohonen ([12]) used it in order to model a claim number processes. In his article, the number of claims is modeled as a mixing Poisson Process, where the mixing random variable follows a shifted Gamma distribution. Willmot and Sundt [16] showed that this Mixing Poisson variable equals to the convolution of a Poisson and a Negative Binomial (a Delaporte) distribution. Ruohonen then argues that the number of claims can be decomposed as the sum of two random variables (a Poisson and a Negative Binomial).

The structure of the article is as follows. Section 2 introduces the model. Maximum Likelihood is suggested in order to estimate the parameters in the model and thus, we provide the appropriate log likelihood function. In Section 3, we discuss how to estimate the inefficiency in the sample. As in the case of Stochastic Frontiers with continuous distributions, the generating process is such that large amounts of heterogeneity in the sample can lead to a non-trivial number of sample points falling underneath the frontier, and thus efficiency measures based on distances from fitted to real values are likely to be inappropriate. As in the article by Jondrow et al [8], we provide an estimator which corresponds to the conditional expected value of the inefficiency term. The evaluation of the small sample performance of the method is discussed in Section 4, and the following section Section [**?**] illustrates the empirical application of the model with a very simple study related to the effect of deprivation on the number of infant deaths in England. Finally Section **??** present some concluding remarks and directions for further research.

2 The Frontier Model

Let us denote by $\mathbf{x}$ the $k \times 1$ vector of covariates which are thought to determine the level of a *technology* frontier, characterized by a count variable. Let us denote the value of the frontier for individual $i = 1, 2, \ldots, n$ in the sample by $Q_i = Q\left(\mathbf{x}_i' \beta\right)$, where β is a vector of unknown scaling constants which must be estimated. The mapping Q is the Discrete (Cost) Frontier providing the minimum expected number of counts given the levels of the covariates $\mathbf{x}$. The frontier quantity Q is unobservable in practice because the discrete frontier is subject to two different sources of variation across sample points. Firstly, as it happens in traditional regression models, Q will be subject to pure stochastic variation, so that the value of Q will oscillate, randomly, above and below the optimal. Therefore, the observed value of the dependent variable is itself a random variable with a, suitable, probability mass function.

Secondly, each sample point will be affected, to some degree, by *technical inefficiency.* The incidence of a technical inefficiency leads to values of the studied discrete variable which are large than the optimal value Q. The magnitude of this increments can itself vary randomly across sample points and, therefore this amount will be represented by a random variable with a probability mass function. Let U denote that additional number of units above Q caused by technical inefficiency. Then, an the observed value of the dependent variable for individual i, denoted Y_i, is given by the following expression:

$$Y_i = Q_i + U_i = Q\left(\mathbf{x}_i'\beta\right) + U_i \tag{1}$$

where $Q : \mathcal{Q} \mapsto \mathbb{Z}^+$ and $U : \mathcal{U} \mapsto \mathbb{Z}^+$ ($\mathbb{Z}^+$ being the union of the set of Natural Numbers and $\{0\}$) are, possibly, correlated random variables.

Given that Y is the sum of two random variables, namely Q and U, its mass function $\mathbb{P}_Y(y) = \mathbb{P}(Y = y)$ can be obtained from standard probability theory which establishes that, when Q and U have range on the set of non-negative integers, the conditional (on $\mathbf{x}$) probability mass function of Y is given by

$$\mathbb{P}_Y(y|\mathbf{x}) = \sum_{q=0}^{y} \mathbb{P}_{Q,U}(q, y-q|\mathbf{x}). \tag{2}$$

where $\mathbb{P}_{Q,U}(.)$ is the joint mass function of Q and U.

Estimation by Maximum Likelihood would be feasible, should we know the actual family of distribution functions to which the joint density $\mathbb{P}_{Q,U}(.)$ belongs. However, from a practical perspective, knowledge of this density will not be available. Fortunately, if we were willing to assume that Q, U are independent random variables, then the above conditional probability would be given by the convolution

$$\begin{aligned} \mathbb{P}_Y(y|\mathbf{x}) = \mathbb{P}_Y(q+u|\mathbf{x}) &= \sum_{u=0}^{y} \mathbb{P}_Q(y-u|\mathbf{x}) * \mathbb{P}_U(u|\mathbf{x}) \\ &= \sum_{q=0}^{y} \mathbb{P}_Q(q|\mathbf{x}) * \mathbb{P}_U(y-q|\mathbf{x}) \end{aligned} \tag{3}$$

which only requires specification of the marginal mass functions of Q and U[1].

2.1 Stochastic Specification under Independence of U and Q

We begin the stochastic specification of the model by characterizing the behaviour of Q, minimal number of counts associated to an experiment. Following a long tradition in statistics and econometrics (see [3] among many others), counts of events can be modeled with the Poisson distribution, and we adopt this probability law here. Therefore,

$$\mathbb{P}_Q(q_i|\mathbf{x}_i) = \begin{cases} \frac{e^{-\lambda_i}\lambda_i^{q_i}}{q_i!} & \text{for} \quad q = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda_i = \lambda_i(\mathbf{x}_i'\beta) = exp(\mathbf{x}_i'\beta) \geq 0$. One important feature of the Poisson Distribution is that (conditional on $\mathbf{x}$) mean and variance have identical value; that is $E[Q_i|\mathbf{x}_i] = var[Q_i|\mathbf{x}_i] = \lambda_i$. This restricts the length of the upper tail of the distribution, so that relatively high values of Q are very rarely attainable. However, experimental data generally exhibits long upper tails (a phenomenon referred to as *overdispersion*). When overdispersion is present in the sample, the Poisson model might be of limited utility. In the setting of our Frontier model, we argue that observed

[1] This assumption is clearly restrictive, although some authors have found it *innocuous* (see [10]) and, probably because of this the assumption underlies most of the literature on stochastic frontier; only recently Smith ([14]) 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.

overdispersion will generally be present, and this is likely to be due to unobserved heterogeneity across individuals. Agents respond differently (and randomly) to the stimulus of $\mathbf{x}$. This different responses will be reflected in Q through λ, the mean value of the distribution, so that heterogeneity in the data can be captured if we allow lambda to vary randomly across observations. Following several precedents in the literature (and the article by Hausman, Hall and Griliches [7] in particular), heterogeneity can be introduced in the model by defining a new mean value parameter as follows:

$$\lambda_i^* = exp\left(\mathbf{x}_i'\beta + u_i\right) = \lambda_i \alpha_i,$$

where $\alpha = exp\left(u\right)$ is a (continuous) random variable such that $\alpha \sim f\left(\alpha\right)$, $E\left(\alpha\right) = 1$ and $var\left(\alpha\right) > 0$. Conditional on α, the density function of Q is

$$\mathbb{P}_Q\left(q_i | \mathbf{x}_i, \alpha_i\right) = \begin{cases} \frac{e^{-\lambda_i^*}\left(\lambda_i^*\right)^{q_i}}{q_i!} & \text{for} \quad q = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

a Poisson Density with parameter $\lambda^\star$, while, integrating out the unobserved α, we obtain

$$\mathbb{P}_Q\left(q_i | \mathbf{x}_i\right) = \int_0^\infty \mathbb{P}\left(q_i | \mathbf{x}_i, \alpha_i\right) f\left(\alpha_i\right) d\alpha_i$$

whose final expression will depend on the particular choice of $f\left(\alpha\right)$. In principle, any continuous distribution on a non-negative support will be suitable; however this choice will have to satisfy a number of conditions. Namely, it must have $range\left(\alpha\right) = \mathbb{R}^+$, it must be adaptable to a wide variety of forms of heterogeneity and, finally, it must yield the mathematical analysis tractable. Following Hausman et al [7] we can let α follow a Gamma distribution with parameters $r = p = \delta$; that is $\alpha \sim G\left(\delta, \delta\right)$. This distribution satisfies the above requirement, and in addition $E\left(\alpha\right) = 1$, $var\left(\alpha\right) = \delta^{-1}$, so that the average expected effect of the heterogeneity in the sample is null (and increasing values of δ reduce the domain of variation of α).

It is well know that the present choice of density for α leads from the Poisson model specified above to a Negative Binomial mass function. The derivation of the result is readily available in a number of sources (see, for example, [6] or [7]), but we reproduce it here for illustrative purposes. Under the Poisson-Gamma model,

$$\begin{aligned} & \int_0^\infty \mathbb{P}_Q\left(q_i | \mathbf{x}_i, \alpha_i\right) f\left(\alpha_i\right) d\alpha_i \\ = & \int_0^\infty \frac{e^{-\lambda_i \alpha_i}\left(\lambda_i \alpha_i\right)^{q_i}}{q_i!} \frac{\delta^\delta}{\Gamma\left(\delta\right)} \alpha_i^{\delta-1} e^{-\delta \alpha_i} d\alpha_i \\ = & \frac{\delta^\delta \lambda^{q_i}}{\Gamma\left(\delta\right) q_i!} \int_0^\infty e^{-\left(\lambda_i + \delta\right)\alpha_i} \alpha_i^{q_i+\delta-1} \left(\frac{\lambda_i + \delta}{\lambda_i + \delta}\right)^{q_i+\delta-1} d\alpha_i \end{aligned}$$

and letting $s = \left(\lambda_i + \delta\right)\alpha_i$,so that $\frac{\partial s}{\partial \alpha_i} = \left(\lambda_i + \delta\right)$,

$$\begin{aligned} \mathbb{P}_Q\left(q_i | \mathbf{x}_i\right) &= \frac{\delta^\delta \lambda^{q_i}}{\Gamma\left(\delta\right) q_i!} \Gamma\left(q_i + \delta\right) \frac{1}{\left(\lambda_i + \delta\right)^{q_i+\delta}} \\ &= \frac{\Gamma\left(q_i + \delta\right)}{\Gamma\left(\delta\right) q_i!} \left(\frac{\delta}{\lambda_i + \delta}\right)^\delta \left(\frac{\lambda_i}{\lambda_i + \delta}\right)^{q_i} \sim \mathcal{NB}\left(p_i, \delta\right) \end{aligned}$$

the expression of a negative binomial density, where in our notation $p = \frac{\delta}{\lambda+\delta}$. It is easy to show that $E(Q|\mathbf{x}) = \delta(1-p)p^{-1} = \lambda$, $var(Q|\mathbf{x}) = \delta(1-p)p^{-2} = \lambda\left(\frac{\lambda}{\delta}+1\right) > E(Q|\mathbf{x})$ and $E[Q|\mathbf{x},\alpha] = \lambda$, $var[Q|\mathbf{x},\alpha] = \lambda(1+\alpha\lambda) > \lambda$ if $\alpha > 0$. Whether conditionally on α or not, the stochastic frontier so defined will admit a longer right tail, and therefore this distribution will be suitable to model those common situations where overdispersion is observed in the data.

Having obtained the conditional density function of Q we can next proceed to complete our frontier model by finding a suitable distribution for the error term U. We note first that if we allow U to follow a Negative Binomial distribution, then the conditional distribution of Y will also be a Negative Binomial distribution; however the parameters of the resulting model will not be identified. There are several feasible choices for this random variable, however a Poisson Distribution seems a logical choice: U must be positive and, since heterogeneity in the sample is assumed to be due to the term α, extreme values of U must be ruled out. Therefore, we advocate that $U \sim \mathcal{P}(\eta)$, where η is the mean inefficiency parameter.

Having completed the stochastic specification of the terms in the model, we finally obtain the conditional density of Y on $\mathbf{x}$, given by the following expression:

$$\begin{aligned}
\mathbb{P}_Y(y_i|\mathbf{x}_i) &= \sum_{q_i=0}^{y_i} \mathbb{P}_Q(q_i|\mathbf{x}_i)\,\mathbb{P}_U(y_i - q_i|\mathbf{x}_i) \\
&= \sum_{q_i=0}^{y_i} \frac{\Gamma(q_i+\delta)}{\Gamma(\delta)\,q_i!}\left(\frac{\delta}{\lambda_i+\delta}\right)^{\delta}\left(\frac{\lambda_i}{\lambda_i+\delta}\right)^{q_i}\frac{e^{-\eta}\eta^{(y_i-q_i)}}{(y_i-q_i)!} \qquad (4)
\end{aligned}$$

The distribution 4 is know as a Delaporte Distribution ([4]), and has been previously employed in actuarial science to model the number of insurance claims against accidents (see [4] and [12]). However, to the best of our knowledge, the genesis of these applications substantially differ from our motivation, and relationship between Delaporte's distribution and stochastic frontier is novel in our analysis. The properties of Delaporte's distribution can be studied generally from the theory pertaining to Mixed Poisson distributions (see Karlis and Xekalaki [9]) since the Delaporte Distribution can be obtained as a mixed Poisson distribution with a shifted Gamma as mixing distribution (Willmot and Sundt [16]). Under this interpretation, Karlis and Xekalaki (pg. 39) give a rule to easily compute the non-central moments of Y as a weighed average of the moments of the mixing distribution. Alternatively note that, under independence, the probability generating function of of Y (given $\mathbf{x}$) equals

$$\phi_{Y|\mathbf{x}}(t) = \phi_{Q|\mathbf{x}}\phi_{U|\mathbf{x}} = e^{-\eta(1-t)}\left(\frac{p}{1-(1-p)e^t}\right)^{\delta} = e^{-\eta(1-t)}\left(\frac{delta}{delta-\lambda(1-t)}\right)^{\delta}$$

where the last equality follows since we defined $p = \delta/(\lambda+\delta)$; replacing t by e^t, we obtain the moment generating function, from which the moments of Delaporte Distribution can be obtained. Clearly, the first and second moments of the Delaporte Distribution can be easily computed given that Y is the sum of two independent random variables. In particular,

$$E(Y|\mathbf{x}) = \lambda + \eta \qquad (5)$$

$$var(Y|\mathbf{x}) = \lambda\left(\frac{\lambda}{\delta}+1\right) + \eta. \qquad (6)$$

That is, by introducing the inefficiency error component in the negative binomial model we cause two effects: firstly, the center of the distribution is shifted by η, the average inefficiency; secondly, the variance of the stochastic frontier is the result of three components (pure stochastic noise, unobserved heterogeneity and inefficiency). By introducing the inefficiency term, our model alleviates the amount of variation encompassed by the Negative Binomial random variable, extracting from the heterogeneity the amount of variation exclusively due to inefficiency.

Finally, the likelihood function of our model, $\mathcal{L}(Y|\beta,\eta,\delta) = \log(\prod_{i=1}^{n} \mathbb{P}(y_i|\mathbf{x}_i))$ equals to

$$\begin{aligned} \mathcal{L}(\mathbf{y}|\beta,\eta,\delta) &= log \prod_{i=1}^{n} \sum_{q_1=0}^{y_i} \frac{\Gamma(q_i+\delta)}{\Gamma(\delta)\, q_i!} \left(\frac{\delta}{\lambda_i+\delta}\right)^{\delta} \left(\frac{\lambda_i}{\lambda_i+\delta}\right)^{q_i} \frac{e^{-\eta}\eta^{(y_i-q_i)}}{(y_i-q_i)!} \\ &= -n\eta + \log(\eta)\, n\bar{y} \\ &+ \sum_{i=1}^{n} \delta \log\left(\frac{\delta}{\delta+\lambda_i}\right) + \sum_{i=1}^{n} \log\left[\sum_{q_i=1}^{y_i} w(q_i)\right] \end{aligned} \tag{7}$$

where

$$w(q_i) = w(q_i|\beta,\eta,\delta) = \frac{\Gamma(q_i+\delta)}{\Gamma(\delta)\Gamma(q_i+1)} \left(\frac{\lambda_i}{\eta(\lambda_i+\delta)}\right)^{q_i} \frac{1}{\Gamma(y_i-q_i+1)} \tag{8}$$

Estimation of the parameters will result from the maximization of the above log likelihood function. Standard statistical theory would then allow us to conclude that the estimator so obtained is consistent for the true parameter values, asymptotically normal and attains the Cramer-Rao lower bound. Test of hypothesis can be implemented based on the classic Wald or Likelihood Ratio Test. In particular, it is of interest to test whether there is overdispersion in the sample, $H_o: \delta = 0$ vs. $H_a: \delta > 0$ and whether there is inefficiency in the sample $H_o: \eta = 0$ vs. $H_a: \eta > 0$. In both situations, the null hypothesis sets the parameter of interest in the boundary of the parameter space, and in these cases the usual critical values from the Chi-Square distribution are not valid (see,for example, Silvapulle and Sen [13]). However, for the Wald test, the asymptotic critical values can be based on those of the one-sided asymptotically normal test.

2.2 Interpretation of the Coefficients

Under the assumptions of the model, and in particular, under the assumption of independence of Q and U, interpretation of the coefficients of the model follows from standard analysis of count data. Thus, let x_j be the j^{th} regressor in the sample. Then the *Marginal Cost* of input x_j is given by $MP_j = \frac{\partial E(Y|\mathbf{x})}{\partial x_j} = \exp(\mathbf{x}'\beta)\beta_j$, so that

$$\beta_j = \frac{\partial E(Y|\mathbf{x})}{\partial x_j} \frac{1}{E(Y|\mathbf{x})}$$

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). Note that the above partial derivative depends on the particular value of the vector $\mathbf{x}$. By calculating this quantity at different values of x_j, we can obtain an *expansion locus* of the Marginal Product of x_j; Alternatively, if a single measure for the partial derivative is required, one could compute the average derivative across the sample: $n^{-1}\sum_{i=1}^{n} \frac{\partial E(Y|\mathbf{x})}{\partial x_{ij}} = n^{-1}\sum_{i=1}^{n} \beta_j \exp(\mathbf{x}_i'\beta)$.

The *Marginal Rate of Substitution* (MRS) of regressor x_j for regressor x_k is defined as the ratio MP_k/MP_j and s measures the rate at which x_j and x_k must be substituted to keep the production function $E(Y|\mathbf{x})$ constant. It follows that $MRTS = \frac{\beta_k}{\beta_j}$, which does not depend on the scale used to measure the regressors.

When any of the regressors enters logarithmically, so that $E(Y|\mathbf{x}) = x_j^{\beta_j}\exp\left(\mathbf{x}'_{-j}\beta_{-j}\right)$ -where the subscript $-j$ indicates removal of element j- then,

$$\frac{\partial E(Y|\mathbf{x})}{\partial x_j} = \beta_j x_j^{-1} E(Y|\mathbf{x}) \Rightarrow \beta_j = \frac{\partial E(Y|\mathbf{x})}{\partial x_j}\frac{x_j}{E(Y|\mathbf{x})}$$

which is the change in the frontier $E(Y|\mathbf{x})$ for a 1% change in x_j.

3 Estimation of Efficiency

After maximum likelihood estimation, information about $(\beta,\eta,\delta)'$ is contained in the vector of ML estimators $(\beta',\eta',\delta')'$. Predicted values of the frontier can be based on $\hat{Q} = \exp\left(\mathbf{x}'\hat{\beta}\right)$. Similarly, *mean inefficiency* can be readily measured with $\hat{\eta}$. However, the ultimate goal of the analysis of technologies is to provide a measure of efficiency for each cross-sectional unit in the sample. From this perspective, $\hat{\eta}$ has very limited scope. Estimations of efficiency cannot be drawn from $Y - \hat{Q}$ either. If the effect of heterogeneity more than offsets the effect of inefficiency, observations will fall below the frontier, and one will end up with an embarrasing negative *residual*. In essence, our problem is similar to that found in traditional stochastic frontiers as in Aigner et al, and not surprisingly, the immediate solution rests on computation of the expected value of U given the information in the sample -as in Jondrow et al [8].

To be more precise, all the information about U is encompassed by Y. All the relevant features of Y are determined by its mass function, which is itself characterized by the vector of parameters $(\beta,\delta,\eta)'$, which can be estimated by consistent methods. Therefore, a way of extracting information about U_i in the sample is to consider

$$E(U_i|Y_i=y_i) = \sum_{u_i=0}^{\infty} u_i \mathbb{P}_{U_i|Y_i}(u_i|y_i) = \sum_{u_i=0}^{y_i} u_i \left\{\frac{\mathbb{P}_{Y,U}(\{Y_i=y_i\}\cap\{U_i=u_i\})}{\mathbb{P}_Y(y_i)}\right\}$$

Clearly, $Y_i = y_i \Leftrightarrow Q_i = y_i - U$, so that

$$\begin{aligned} E(U_i|Y_i=y_i) &= \sum_{u_i=0}^{y_i} u_i \left\{\frac{\mathbb{P}_{Y,U}(\{Q_i=y_i-u_i\}\cap\{U_i=u_i\})}{\mathbb{P}_Y(y_i)}\right\} \\ &= \sum_{u_i=0}^{y_i} u_i \left\{\frac{\mathbb{P}_Q(y_i-u_i)\,\mathbb{P}_U(u_i)}{\mathbb{P}_Y(y_i)}\right\} \\ &= \sum_{u_i=0}^{y_i} u_i w_i \end{aligned} \tag{9}$$

a weighted average resulting from equations (3) and (4) and where the weights equal to

$$w_i = \frac{\frac{\Gamma(y_i-u_i+\delta)}{\Gamma(\delta)(y_i-u_i)!}\left(\frac{\delta}{\lambda_i+\delta}\right)^{\delta}\left(\frac{\lambda_i}{\lambda_i+\delta}\right)^{y_i-u_i}\frac{e^{-\eta}\eta^{(u_i)}}{u_i!}}{\sum_{u_i=0}^{y_i}\frac{\Gamma(y_i-u_i+\delta)}{\Gamma(\delta)(y_i-u_i)!}\left(\frac{\delta}{\lambda_i+\delta}\right)^{\delta}\left(\frac{\lambda_i}{\lambda_i+\delta}\right)^{y_i-u_i}\frac{e^{-\eta}\eta^{(u_i)}}{u_i!}} \tag{10}$$

A feasible estimator of the inefficiency can be based on

$$U_i^* = \sum_{u_i=0}^{y_i} u_i \hat{w}_i$$

where $\hat{w_i}$ equals w_i but with β, η and δ replaced by their Maximum Likelihood counterpart. Note, once again, that this estimator will return the expected inefficiency given the information available in the sample; as the expression of the estimator itself shows, we are smoothing across the feasible values of the inefficiency term, and hence we do not expect to obtain the actual value of the inefficiency. Secondly, this estimator does not produced integer valued estimations, but a vector of real numbers; of course, U could then be approximated by taking the integer part of the the produced estimation. In this case, *intU* may be understood as a class of inefficiency, so that all observations with $U^* = u^*$ would belong to such a class.

4 Monte Carlo

To study the finite sample behaviour of the model, we conducted two Monte Carlo experiments. Version V.4.0.4 of Ox was used in order to generate observations from the following model:

$$Y_i = Q_i + U_i \text{ for } \begin{cases} Q_i \sim \mathcal{P}\left(\lambda_i^\star\right) & \lambda_i^\star = \lambda\left(\mathbf{x}_i'\beta\right)\alpha_i \\ U_i \sim \mathcal{P}\left(\eta\right) & \\ \alpha_i \sim \Gamma\left(\delta, \delta\right) & \end{cases}$$

where $\mathcal{P}$ is a Poisson Distribution with mean represented by the term inside the brackets, Γ denotes the gamma distribution and $\mathbf{x}'\beta = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i}$. Values of X_{ji} where drawn from uniform distributions on the interval $[0, 5]$.

In our fist study, we analyze the relative merits of the model under a variety of data generating processes. We fixed the valued of the parameters in the structural part of the frontier at $\beta_o = 0.5$, $\beta_1 = 0.3$ and $\beta_2 = 0.4$, while different combinations of δ and η were used in order control the amount of heterogeneity and inefficiency in the sample. Small values of δ (so that $var\left(\alpha\right)$ is large) and η leads to samples where heterogeneity dominates inefficiency, while increasing the value of δ permits U, the inefficiency or *technical hazard* term to play a more significant role. The sample size was kept constant at 5000 observation.

Table 1 presents the results obtained after applying our model to six different generation processes. The first three models have common value of $\delta = 0.5$, while the increasing value of the parameter η represents increased levels of inefficiency in the sample. The last three models have fixed (low) levels of inefficiency ($\eta = 1$) and increasing levels of heterogeneity.

Generally speaking, the method performs well, and we obtained estimated values close to the true parameter values. Estimation of the interception, however, is subject to downward biases, which in cases can be rather large (as it is the case with Model 4). However, a second experiment was designed in order to shed some light on the consistency properties of the estimator, which we conjectured would follow from standard Maximum Likelihood theory. We drew samples of varying size (from 100 to 3200 observations), from a model such that $\beta_o = 0.2$, $\beta_1 = 0.2$, $\beta_2 = 0.3$ $\eta = 2$ and $\delta = 1$ (corresponding to Model 6 in the previous experiment). Each sample of size n was drawn 1000 times, and the sample Mean Squared Error of was computed for each of the parameters in the model. A cross section of the results of this experiment are collected in Table 2. With a sample of

Monte Carlo Experiment 1
$(\beta_o = 0.5, \beta_1 = 0.3, \beta_2 = 0.4)$
Inefficiency/Heterogeneity In Brackets (η, δ^{-1})
Sample Size $N = 5000$

	Model 1 $(.5, 2)$	**Model 2** $(4, 2)$	**Model 3** $(8, 2)$	**Model 4** $(1, 0.25)$	**Model 5** $(1, 1)$	**Model 6** $(1, 2)$
η	0.43834	3.9250	8.1271	1.4903	0.96096	0.95953
	(0.043610)	(0.092292)	(0.1133)	(0.16310)	(0.08181)	(0.056107)
δ^{-1}	1.8729	1.8718	2.1574	0.27246	1.0279	1.8899
	(0.057140)	(0.071097)	(0.088935)	(0.012275)	(0.036168)	(0.060773)
β_o	0.49939	0.45901	0.42819	0.38523	0.45042	0.51320
	(0.056473)	(0.064216)	(0.069856)	(0.06192)	(0.050528)	(0.057962)
β_1	0.29158	0.31488	0.30428	0.31349	0.30820	0.29619
	(0.013812)	(0.014559)	(0.015932)	(0.0085434)	(0.011072)	(0.014202)
β_2	0.40775	0.40894	0.42704	0.41495	0.40933	0.40815
	(0.014125)	(0.014886)	(0.015795)	(0.010062)	(0.011317)	(0.014253)

Table 1: *Estimated parameters for a variety of models.*β_0, β_1 *and* β_2 *were kept fixed at 0.5, 0.3 and 0.4 respectively, while* η, δ *varied in accordance to the patter explained under each Model's heading.*

just 1000 observations, the MSE of all the parameters in the model is relatively low (although there is, of course, margin for improvement). The MSE decreases rapidly as the sample size doubles and, with sample sizes of 400 observations, one already obtains dramatic gains in the accuracy of the method. Our simulations show that with samples of 2 or 3 thousand observations small MSE are expected, and therefore, this seems to confirm that the property of consistency is inherited from the Maximum Likelihood method.

Overall, this simulation reveals that the proposed model is flexible, adapting to a wide variety of generating process, while it exhibits a good performance even with relatively small sample and, although there is margin for improvement, the results here obtained are similar to those reported in many other articles introducing Maximum Likelihood Models.

5 Application: Infant Mortality and Income Deprivation

This section illustrates the empirical implementation of the model. We shall undertake a small study regarding the potential relationship between the levels of deprivation the number of infant deaths. Our study does not intend to be rigorous but illustrative of the type of problem our model could be applied to. The final objective is to isolate the effect of environmental and economic deprivation on the observed count of deaths.

Our data set was extracted from the website of the UK Office of National Statistics. Available

Monte Carlo Experiment 2
Mean Square Error
Model Specification: $(\eta, \delta, \beta_o, \beta_1, \beta_2) = (2, 1, 0.2, 0.2, 0.3)$

	Parameter				
N	η	δ	β_o	β_1	β_2
100	0.49443	0.18699	0.28340	0.0089427	0.10350
200	0.29158	0.094135	0.14647	0.0041765	0.0054366
400	0.14302	0.048145	0.070838	0.0020032	0.0025
800	0.068315	0.021977	0.031249	0.00092852	0.0011795
1600	0.033138	0.011838	0.016354	0.00046541	0.00058106

Table 2: *Mean Square Error*

to us were the counts of infant deaths in the UK during 2001 as well as a number of environmental, social and economic variables. All the variables refer to Local Authorities (354 of them in England).

By definition infant deaths are deaths at ages under one year old. It is conjectured that the mortality frontier would depend on a number of social, economic and environmental aspects of the region. Firstly, the number of infant deaths will be correlated with the number of births at a given area, itself determined by the size of the population in the area. We shall denote by (p_i) the natural logarithm of the number of inhabitants in area i, in accordance to the 2001 Census.

Environmental factors have been also associated to the overall health of the population (see, for example, Tulchinsky and Varavikova [15], 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 general a by product of fuel consumption (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. In our sample we collect the effects of these gases. by introducing the logarithm of the Local Area Average of NO_x emissions intensity score, denoted NO_i^s. 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.

In order to capture the levels of economic deprivation, the logarithm of the percentage of active population claiming unemployment benefits, u_i is introduced in order to capture levels and potential impact of economic deprivation. One of the risk factors associated to infant mortality is the baby's weight at birth. Low birthweight (less than 2.5Kg) is an indicator of newborn's chances of survival (and later development in life); babies who were not properly fed during in the womb, face a higher risk of dying in the earlier years of life, and thus, they are likelier to die in the first 12 months of life. Therefore, our model includes the logarithm of regional percentage of underweight newborns in 2001, denoted (w_i). Note that NO emissions and economic deprivation are likely to affect the proportion of underweight babies in a given area; hence, the estimated coefficient of the log percentage of underweight newborns is likely to capture the effect of the environmental and economic deprivation on the proportion of underweight babies, as well as a measure of the proportion of underweight newborns who die within a year. Thus, we conjecture that after including w_i in the model, the estimated coefficient of NO_i^s and u_i should then measure the direct net effect of these variables on

Estimated Model
$N = 354$

Parameter	Estimate	Std. Err.	t-statistic	MP	Poisson	NB
η	1.4482	0.79417	1.8236	$--$	$--$	$--$
$1/\delta$	0.045091	0.0077396	5.8260	$--$	$--$	0.04158
β_o	-12.350	0.77290	15.979	$--$	-11.50016	-11.24123
Population	1.1161	0.057918	19.270	26.313	1.043059	1.046322
Weight	0.41722	0.11540	3.6156	9.8366	0.5555207	0.3501964
NO	0.40408	0.063642	6.3493	9.5268	0.3780584	0.356472
Benefits	0.32103	0.064347	4.9891	7.5687	0.2664801	0.321683

Table 3: Estimated Frontier

number of deaths.

The frontier model is therefore:

$$\lambda_i = \exp\left(\beta_o + \beta_1 p_i + \beta_2 w_i + \beta_3 NO_i^s + \beta_4 u_i\right)$$

The results are summarized in table 5 -results from Poisson and Negative Binomial fits are also reproduced for comparison. The estimator of the expected inefficiency, 3, was computed for each sample point. An estimator of the density of U^* is given in Figure 1. The range of values we obtained was $[0, 2.6042]$, with an average value of 1.34, naturally close to the estimated value of eta. The estimated standard deviation equals 0.29. The estimated coefficients of skewness and kurtosis are 0.23 and 3.611 which are close to the values of the normal distribution. The mode of the distribution is located at 1.19, while the median is at 1.33; the 75 percentile is located at 1.48.

The first aspect of interest in this type of study would be the levels of inefficiency and heterogeneity in the sample. In accordance to the estimations, the average estimated inefficiency is 1.4482, while the estimated value of the variance of the heterogeneity term δ^{-1} is 0.045 which implies the existence of moderate to low levels of heterogeneity in the sample. It seems as if discrepancies across Local Authorities is mostly explained by inefficiency. Among the regressors, the percentage of underweight babies, emissions of nitrogen oxides and economic deprivation are all statistically significant in order to explain the absolute levels of infant mortality. Their positive coefficients imply that all these are risk factors (as expected) and higher levels of any of these variables leads to higher levels of expected deaths. In order to estimate the magnitude of expected change on the expected count of deaths the average estimated marginal costs associated to each of the components is calculated, and the results are shown in column 4 of the table. Ignoring the effect of w_i for the reasons outlined above, the most relevant risk factor is the level of NO emissions rather than levels of economic deprivation. The marginal rate of substitution of NO emissions with respect to the log count of underweight babies is of 1.2587, which would imply that policies oriented to cutting emissions might contribute more effectively to reduction of deaths than policies oriented toward reducing economic deprivation.

6 Conclusions

This article is a first attempt to construct a maximum likelihood based model for Stochastic (Cost) Frontier Models when the dependent variable is a count. This type of model arises naturally in a

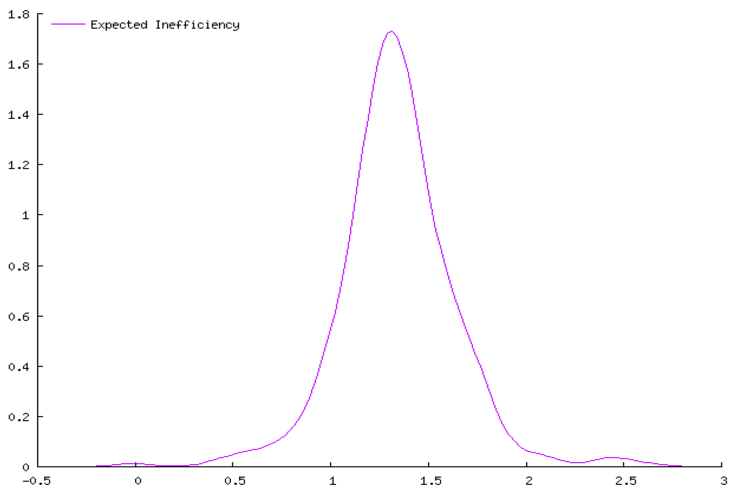

Figure 1: Density Estimator of $U^* = E(U|Y)$.

number of statistical scenarios, including Epidemiological studies. Our model, which inherits the theoretical properties of maximum likelihood estimators, is capable of accurately estimating the frontier function, and at the same time it effectively separates heterogeneity and inefficiency from pure statistical noise. We also propose a measure of inefficiency. As in the paper by Jondrow et al, our measure is based on the expected value of the inefficiency term, and it is therefore a smooth function of the information in the sample. In our case, therefore, estimated inefficiency must be understood more as a rank than as the actual value of the inefficiency in the sample. Alternative model specifications to ours are possible; for example, one could have argued that the frontier is likely to be common to all individuals in the sample, while heterogeneity would me more likely to affect the inefficiency term. In such a case a more suitable specification would be a Poisson-Negative Binomial, the latter characterizing the inefficiency term. One could, more generally, explore other combinations of mixture models to allow heterogeneous frontiers and inefficiencies, but this is likely to complicate the analysis, and therefore we leave it for future research.

References

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

[2] Cameron, C., and Trivedi, P. Econometric models based on count data: Comparison and applications of some estimators. *Journal of Applied Econometrics 1* (1986), 29–53.

[3] Cameron, C., and Trivedi, P. *Regression analysis of count data.* Cambridge University Press, 1998.

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

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

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

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

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

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

[10] Kumbhakar, S. C., and Knox Lovell, C. *Stochastic Frontier Analysis.* Cambridge University Press, 2003.

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

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

[13] Silvapulle, M., and Sen, P. *Constrained Statistical Inference.* Wiley, 2006.

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

[15] Tulchinsky, T., and Varavikova, E. *The New Public Health.* Academic Press, 2000.

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