The University of Manchester

# Economics
# Discussion Paper Series
# EDP-0916

# Count Data Stochastic Frontier Models, with an application to the patents-R&D Relationship

Eduardo Fé-Rodríguez
Richard Hofler

September 2009

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

# Count Data Stochastic Frontier Models, with an application to the patents-R&D Relationship


Eduardo Fé-Rodríguez*
*University of Manchester*

Richard Hofler †
*University of Central Florida*


September 9, 2009


**Abstract**

This article introduces a class of Count Data Stochastic Frontier models which researchers can use in order to study efficiency in production when the output variable is a count (so that its conditional distribution is discrete). Two members in this class are studied at length. The first model is suitable whenever unobserved cross-sectional heterogeneity is not likely to be a problem. Conversely, for those cases when it must not be ignored, we propose a second model which takes into account the potential effect of cross-sectional unobserved heterogeneity. A Monte Carlo study is presented in order to evaluate the merits of these two models in small samples. Finally, a new approach is proposed to study the relationship between R&D investment and patents at the firm level. We suggest estimation of the underlying production frontier of patents via the Count Data Stochastic Frontier models proposed here, unlike in previous research based the estimation of the conditional mean of a count. Estimates of elasticity of patents and the average efficiency in the production of patents are provided.



**Key Words**: *Count Data, Stochastic Frontier, Efficiency, Distance Function, Heterogeneity*

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



*Health Economics. Room 1.308, Jean McFarlane Building. University of Manchester. Manchester, M13 9PL, UK. email:e.fe-rodriguez@manchester.ac.uk.

†Department of Economics, University of Central Florida. Orlando, FL, 32816-1400. email: rhofler@bus.ucf.edu

## 1 Introduction

There is a thriving literature about the estimation of Stochastic Frontiers and average efficiency scores which has origin in the seminal article by Aigner et al. (1977) (ALS in what follows). Early contributions developing these models further are Stevenson (1980), Schmidt and Sickles (1984), Greene (1990), Cornwell et al. (1990) and Battese and Coelli (1995). The fully parametric flavour of these articles has been recently attenuated with the intention of reducing the chances of model misspecification. Thus Park et al. (1998), Sickles (2005) and Park et al. (2007) explore the literature on semiparametric efficiency bounds in order to propose a efficient panel data method; Martins-Filho and Yao (2007) present a fully nonparametric estimator of a frontier, while Kumbhakar et al. (2007) use the local nonlinear least squares method of Gozalo and Linton (2000) in order to propose a highly adaptable stochastic frontier model. All of these (and other) estimators and models have been developed with a continuous dependent variable in mind. But there are situations when the output variable of interest is a count (for example, the number of patents obtained by a firm or the number of infant deaths in a region) and the mean of this count is low. It is natural that a researcher facing such a count dependent variable case would desire to maintain the discrete aspects of production within the analysis, and therefore the question is whether traditional continuous data methods (with base on Aigner et al. (1977) and Schmidt and Sickles (1984)) are still suitable in this setting. We find three reasons why a new model is advisable.

The first argument is implicit in the count data literature. It has been known for some time that approximating the distribution of a discrete random variable by that of a continuous random variable can lead to deficiencies (in terms of the quality of the resulting estimates), since this might represent one of those model misspecifications discussed in the important article by White (1982) (for comments on this issue see, for example, Hausman et al. (1984) and the comprehensive reviews by Cameron and Trivedi (1998) and Winkelmann (2008)). These deficiencies have been the driver of much of the literature on mean regression of count data and yet, not much work has been devoted to develop a Count Data Stochastic Frontier model (CDSF hereafter). Therefore, in order to avoid model misspecifications resulting in inefficient and, probably, inconsistent estimates, a method supported by discrete distributions is due now.

The second argument is of theoretical nature. As we discuss below, underlying SFM there is a multiplicative distance function relating output ($y$), frontier ($f$) and inefficiency ($0 < d < 1$), so that $y = fd$. In applications, it is customary to transform the model into logarithms, so that $\log y = \log f + \log d$ (noting that $\log d < 0$). However if $y$, $f$ and $d$ are discrete valued, the multiplicative scheme is unlikely to be satisfied. For instace, let $y = 7$ and $f = 3$; then there is not integer value for $d$ solving assumed identity. This advices the contruction of a new distance function suitable for count data.

The final argument, of empirical nature (but naturally related to the later), is the non-negligible frequency with which zero values of the count dependent variable are observed. This is a feature pervading most count data applications, and difficulting drawing comparisons of goodness of fit between models with continuous distributions and those based on a discrete probability distribution. It also has a significant impact on the applications we have in mind. Firstly, 0 production with positive $f$ is a perfectly reasonable scenario, but unfeasible with a multiplicative distance function unless $f = d = 0$. Secondly, and more importantly, $\log 0$ is not defined and so the logarithmic transformation of the multiplicative distance function is not defined in such cases. Since applications hinge on such a transformation, one is left with situations when the method is undefined for a significant proportion of the sample. Although one could probably try to circumvent these problems by transforming these 0 observations into a continuous random variable (by adding a random non-negative, but arbitrarily small, number), this does not seem a satisfactory strategy, especially when the number of 0 in the sample is substantial.

To the best of our knowledge, there are only two articles exploring CDSF to date. The first reference is Fé-Rodríguez (2008)[1] who explores the estimation of a frontier for economic bads when output and inefficiency are discrete valued random variables. The second reference is Hofler and Scrogin (2008). These authors note that the actual effect of inefficiency on production is equivalent to an underreporting of the final output, and therefore the Beta-Binomial model for underreported counts (see, for instance, Fader and Hardie (2000)) is suggested to approximate the production function of an economic good. Both models are based, in effect, on a Mixed Poisson distribution, but while Hofler and Scrogin (2008) assume that inefficiency is continuously distributed, Fé-Rodríguez (2008) assumes that output and inefficiency enter the model as count variables. Apart from incorporating into the analysis the count nature of the data, the strength of these models is that they can treat unobserved cross-sectional heterogeneity and inefficiency in production as two separate sources of variation in the data. This is a rare property in the literature of Stochastic Frontiers, the exception being Greene (2005). This author pointed out that, in the context of a fully specified model of production, cross-sectional heterogeneity refers to factors, other than inefficiency, which are directly observable but explain part of the variation in the sample. Therefore omitting this variable from the model will bias the estimates in a generally unpredictable way.

Despite their advantages, the models in Fé-Rodríguez (2008) and Hofler and Scrogin (2008) are designed with either an economic good or an economic bad in mind, and they cannot be used to analyse both types of commodities. This is seemingly irrelevant since a researcher could use one method or the other depending on the nature of the commodity under study. However, it is implicit in the literature that a proper Stochastic Frontier estimator must be adaptable to commodities of an arbitrary nature so that a simple transformation within the model (normally a change of sign) is sufficient to capture the nature of the commodity whose frontier one needs to estimate. From this perspective, neither Fé-Rodríguez (2008) nor Hofler present proper Stochastic Frontier models. Therefore, the purpose of this paper is to provide a proper Count Data Stochastic Frontier in such a way that the discrete nature of output is preserved.

In section 2, we present the model in its most general form. The starting point is

[1] An earlier version of this paper is Fé-Rodríguez (2007), which was presented at European Workshop on Efficiency and Productivity Analysis V, Lille, France, 2007.

the definition of a distance function, suitable for count data, which allows us to progress toward an econometric CDSF model. The distance function here provided naturally leads to a Mixed Poisson regression, where the mean parameter happens to be the frontier of production, while inefficiency appears as an additive random effect.

In section 3, the half normal distribution is suggested as a candidate distribution to model inefficiency, as it is an adaptable distribution, meaningful from an economic point of view and which leads to a conditional distribution of output with well defined moments. This last point is not a trivial one. As we discuss in this article, when output is an economic bad, other more attractive distributional choices for the random effect used to model inefficiency (like for example the Gamma distribution) lead to conditional distributions which have no lower order moments. We will say that these other models do not treat goods and bads symmetrically and, although applicable for efficiency analysis in the production of an economic good, they are discarded from consideration in this article.

Section 4 extends the model in order to account for unobserved cross-sectional heterogeneity, and then the performance of both models is evaluated in a limited Monte Carlo experiment, in Section 6. Section 5 discusses how to produce cross-sectional estimates of the levels of attained efficiency. Our measure of efficiency is based on the posterior expected value of the random effect capturing inefficiency -in a fashion similar to Jondrow et al. (1982)- and it produces unbiased estimates of the average efficiency level in the sample. It has been pointed out by Wang and Schmidt (2009) that this statistic can be used to estimate the efficiency level given the distributional choices for inefficiency and output, and therefore, the estimated distribution arising from our statistics need not to be confounded with the actual (unconditional) distribution of inefficiency. Furthermore, estimates like the ones here presented, based on the Jondrow et al. (1982) methodology, are a shrinkage toward the mean, so that the estimated distribution will exhibit a narrower domain than the actual distribution. Although this does not seem to imply a problem from an empirical perspective (as the estimate of the average efficiency is still unbiased), researchers must be aware that if inference is to be based on the estimated distribution of efficiency, comparisons must be done with the distribution of estimated efficiency conditional on the distribution of actual efficiency, rather than just the latter.

This paper presents an application of our methods to the study of the relationship between expenditure on research and development (R&D in what follows) and the number of patents at firm level. This relationship has attracted the attention of economists because patents may be taken to represent the value of the underlying stock of knowledge in a firm, and therefore knowledge of the production function of patents can help to estimate important policy measures such as the elasticity of R&D. Previous analyses have put emphasis on approximating the production function of patents via mean regression models; however this method does not take into account the fact that some firms will be more efficient than others in handling their expenditure on R&D. Furthermore, it seems intuitively clear that whatever happens at the mean of the distribution might differ substantially from the behaviour of the distribution in the upper quantiles, where the actual frontier of production possibilities is likelier to be. Therefore, we try to capture that behaviour via our CDSF. Our results are presented in Section 7, and Section 8 presents some concluding remarks.

## 2 A model of production in discrete amounts

Consider a collection of competing firms or institutions ($i = 1, \ldots, n$) producing a commodity in quantities $y_i$, and suppose that we are interested in the behaviour of output and efficiency in production, conditional on a number of observed variables. The commodity under study could be an economic good (for example, we could be researching the number of patents awarded to firms), but it could also be an economic bad where, for instance, the focus of the research could be the number of infant deaths observed in a group of geographical units. In the first case, agents engaged in production try to maximize output, while in the second instance authorities will want to cap the incidence of the final product. However, and unlike in the typical application, the dependent variable of interest, $y$, is a count and its average value is typically low.

### Distance Functions

It is well known that underlying ALS-type models there is an implicit distance function capturing the discrepancies between ideal and actual outputs (see, for example Kumbhakar and Lovell (2003)). For continuous data and in the case of, for instance, the production of an economic bad, this distance function is defined as:

$$\varepsilon(x, y) = \sup_{\lambda} \{y/\lambda \geq h(x)\} \tag{2.1}$$

where $h(.)$ is the production function of a commodity $y$, so that $1 \leq \varepsilon < \infty$ measures the discrepancy between actual output, $y$, and ideal output, $h(.)$, given input levels $x$. From this distance function it follows that $y = f(x)\varepsilon(x, y)$, or equivalently $\log y = \log(h(x)) + \log(\varepsilon(x, y))$. Aigner et al. (1977) build an econometric model over the last equality, by letting $\varepsilon(x, y) = \varepsilon$ be a random variable with domain on the positive real line, and then assuming that $\log(h(x; \beta))$ (a parameterization of $h(x)$) is the conditional mean of a symmetric-about-zero random variable, (where $\beta$ is a conformable vector of scaling constants). A very remarkable feature of this model is that a model for economic goods can be generated by a simple change of sign in the decomposition of output (so that $\log y = \log(h(x)) - \log(\varepsilon(x, y)))$, and both ensuing models are well defined and have, at least, finite first moments.

In the framework of discrete valued data, it could be argued that both the production frontier and the variable collecting inefficiency in production ought to be discrete valued (as the output variable is). However, when $h(x)$, $\varepsilon(x, y)$ and $y$ are all discrete valued, the multiplicative scheme underlying ALS rarely will hold (e.g. the point $(h, \varepsilon, y) = (7, 4, 3)$ is a sensible hypothetical description of production, and yet not admissible through a multiplicative scheme, since $4 \times 3 \neq 7$). Another practical problem arises because when $y$ is discretely distributed the probability of observing the value 0 is non-negligible (and the magnitude of this probability can be substantial). Then, $\log(y)$ is not defined, and the above distance function is not applicable. It is therefore customary to introduce a new distance function taking the discrete nature of the data into account.

In the context of production of economic bads, Fé-Rodríguez (2008) proposed the following distance function:

$$\varepsilon^*(x,y) = \max_{l\in\mathbb{N}} \{y - l \geq h(x)\} \Rightarrow y = h(x) + \varepsilon^*(x,y). \tag{2.2}$$

where now $y$, $h(.)$ and $\varepsilon(.)$ are discrete valued and output is naturally described as the convolution of the levels of the production frontier with the level of inefficiency. Econometric modeling is then straightforward: one only needs to assign discrete probability distributions to each term, and estimate the resulting convolution by some suitable method (often Maximum Likelihood).

This approach, however, is not adaptable to economic goods and bads alike since, unlike with the distance function underlying ALS, a simple change of sign in the right hand side cannot be used to change the focus from production of economic bads into production of economic goods. The reason is that by doing so, one could encounter situations when the distributional choices for $h(.)$ and $\varepsilon(.)$, although suitable for economic bads, do not preclude negative counts of the output variable in the context of economic goods, unless enough regularity is imposed to ensure that $h(x) - \varepsilon^*(x,y) \geq 0$. This is at odds with the assumption of non-negative production.

### A New Distance Function

Since none of the above approaches seems satisfactory from an econometric perspective, in this article we look for a different, more suitable distance function for a proper CDSF. This is achieved by allowing inefficiency to be continuously distributed, but maintaining the discrete nature of the output variable. For simplicity, the discussion that follows focuses on the case of economic goods (the discussion extends to economic bads immediately). Consider again the multiplicative distance function underlying ALS:

$$u^\star(x,y^\star) = \inf_{u\in[0,1]} \left\{\frac{y^\star}{u} \leq h(x)\right\} \Rightarrow y^\star = h(x)u^\star(x,y) \tag{2.3}$$

In the latest scheme none of the quantities $h(.)$, $u^\star(.)$ or $y^\star$ are assumed to be discrete valued, and only $u^\star(.)$ is restricted to take values on $[0,1]$ (in the case of economic goods). This is in contrast with the approach in Fé-Rodríguez (2008) (who assumes an additive distance function with $y$, $h(.)$ and $\varepsilon(.)$ all discrete valued) and it is also in contrast with ALS in two ways. Firstly, unlike ALS, we do not need to transform the distance function by taking logarithms (and yet, some regressors could have been transformed that way). Secondly, it is now assumed that $y^\star$ is not observed. Instead, we observe the discrete quantity $y = 0, 1, \ldots$, which is related to $y^\star$ in such a way that, $y = j$ (a non-negative integer) with a prescribed probability whenever $y^\star$ attains a level, say, $(a,b] \subseteq \mathbb{R}$.

The econometric equivalent of (2.3) can then be constructed as follows. Firstly, as in ALS we assume that inefficiency is directly unobservable. This could be specified as $u = \exp(-\varepsilon(x,y)) = \exp(-\varepsilon)$ where $\varepsilon$ behaves as a random variable with domain on $\mathbb{R}^+$ and with density function $f(\varepsilon;\theta)$, for some vector of parameters $\theta$. Similarly, $h(x) = \exp(x'\beta)$ (deterministic) for some vector of parameters[2] $\beta$. This defines $y^\star$ as a random variable on $\mathbb{R}^+$. Secondly, the observed value $y$ depends on the levels attained by $y^\star$ but not in a deterministic way; because of unexpected contingencies of all kinds, we assume that for given $y^\star$ the observed output variable takes values on $\mathbb{N}$ with a prescribed probability. In particular, assume that, conditional on $\varepsilon$, this probability law is a Poisson distribution, with mass function

$$\mathbb{P}(y_i|u_i, x_i) = \frac{e^{-\tilde{\lambda}_i}(\tilde{\lambda}_i)^{y_i}}{y_i!} \tag{2.4}$$

where $\tilde{\lambda} = \lambda u = \exp(x_i'\beta - \varepsilon)$, which is our Count Data model for economic goods.

More generally, we define the class of Count Data Stochastic Frontier (CDSF) models as those based on the distance function (2.3) -and its equivalent for economic bads- with mass function:

$$\mathbb{P}(y_i|x_i, u_i) = \begin{cases} \frac{e^{-\lambda_i u_i}(\lambda u_i)^{y_i}}{y_i!} & \text{for } y = 0, 1, \dots \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

where $\lambda_i u_i = \exp(x'\beta \pm \varepsilon_i)$ and $\varepsilon$ is assumed to follow a certain distribution.

# 3 Count Data Stochastic Frontier Models (CDSF)

In order to obtain a functioning CDSF we need to select a distribution for the term $u = \exp(\pm\varepsilon)$. There are potentially as many different models for the conditional distribution of output as there are distributions to choose for $\varepsilon$. However, our choice should result in a model for $y$ with a tractable conditional distribution for the output count variable and whose moments are well defined irrespectively of the nature of the commodity under analysis. Tractability must be understood in a broad sense. Although estimation via Maximum Likelihood and Method of Moments has traditionally been based on relatively simple, closed-form distributions and moments, there is a vast literature on simulation techniques, such as Gaussian Quadrature (see, for instance, Stroud (1971), Butler and Moffitt (1982), Greene (1990) or, Winkelmann (2008)) or Pseudo Monte Carlo integration (McFadden (1989), Geweke (1995), Greene (2003), or Gentle (2003)) which allows estimation of parameters in highly flexible, albeit otherwise intractable, models. The existence or lack of moments is the key property that will allow us to make informed decisions regarding what distribution to attach to $\varepsilon$, the random variable capturing inefficiency: as we discussed below, some popular distributional choices lead to models whose lower order moments might not exist when $y$ is an economic bad. In this sense, these models treat good and bads asymmetrically.

[2]The choice of exp(.) as link function is only for convenience, and it ensures that $h(.)$ and $u$ are non-negative. Other functions are available. Similarly the choice of a single index function for $h(.)$ is for convenience, as nothing precludes more involved mappings relating $x$ with $y^*$

In the search for a suitable distribution for $\varepsilon$, our attention was drawn to the half normal distribution, whose density function is

$$f(\varepsilon; \sigma_\varepsilon) = \frac{2}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}} \mathbb{I}_{[0,\infty)} \text{ for } \sigma_\varepsilon > 0 \tag{3.1}$$

This distribution is attractive for a number of reasons. It is moderately adaptable, thanks to its scale parameter and the varying shape of its density function. Its density exhibits varying returns to efficiency: the probability of observing inefficiency levels away from zero has increasing marginal products, up to an inflection point beyond which the marginal products are still negative but increasing. Finally, this choice will render a tractable conditional distribution for $y$ that has, at least, well defined first two moments.

With this choice for $f(\varepsilon; \sigma_\varepsilon)$, the conditional distribution of $y$ is defined as follows:

$$\begin{aligned} &\mathbb{P}(y|x;\theta) \\ =\ &\int_0^\infty \frac{(\exp(-\exp(x'\beta \pm \varepsilon))) \exp(y(x'\beta \pm \varepsilon))}{y!} \frac{2}{\sigma_\varepsilon \sqrt{2\pi}} \exp(-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}) d\varepsilon \\ =\ &\int_0^\infty \frac{(\exp(-\exp(x'\beta \pm \epsilon\sigma_\varepsilon\sqrt{2}))) \exp(y(x'\beta \pm \epsilon\sigma_\varepsilon\sqrt{2}))}{y!} \frac{2}{\sqrt{\pi}} \exp(-\epsilon^2) d\epsilon \end{aligned} \tag{3.2}$$

for $\theta' = (\beta', \sigma_\varepsilon)$. The change of variable $\varepsilon = \epsilon\sigma\sqrt{2}$ led to the second equality. After some work, it soon becomes apparent that if $\varepsilon$ is distributed half normal, then one cannot obtain a closed form expression for the ensuing distribution of $y$. However, this expression is an integral of the type computable -up to a small error- via Gauss-Hermite quadrature (see Press et al. (1992) or Judd (1998)).

Gaussian Quadrature methods are based on the generation of a set of nodes and weights which are optimal, so that the integral is accurately approximated. Quadrature is, in this sense, a deterministic method unlike Monte Carlo integration, and it is not as demanding in terms of computer power as (Pseudo) Monte Carlo approaches since only a very small number of points are required in order to obtain accurate approximations of the underlying integral (with only 10 points can often obtain great accuracy). If we adopt this technique here, the conditional distribution of $y$ can the be approximated by the sum:

$$\begin{aligned} &\mathbb{P}(y|x;\theta) \approx \mathbb{P}(y|x;\theta;\xi_j) \\ =\ &\frac{2}{\sqrt{\pi}} \sum_{j=1}^{J} \left[ \frac{(\exp(-\exp(x'\beta \pm \xi_j\sigma_\varepsilon\sqrt{2}))) \exp(y(x'\beta \pm \xi_j\sigma_\varepsilon\sqrt{2}))}{y!} \right] w_j \end{aligned} \tag{3.3}$$

where $\xi_j$ are the $J$ nodes (or points in the domain) at which the kernel of the integral in (3.2) is evaluated, and $w_j$ are the corresponding values of the weighting function $\exp(-\epsilon^2)d\epsilon$. It is customary to note that, because the underlying integral runs on $[0, \infty)$ (that is, we are assuming the integrand in (3.2) is 0 on $\mathbb{R}^-$) only the positive nodes of Gauss-Hermite quadrature need to be used in the evaluation. Although there is nothing wrong with that method, an alternative approach which employs all positive and negative points of quadrature can be devised. The key step requires noting that, in general, for some deterministic function $g(.)$, the following identity holds:

$$\mathbb{P}(y|x;\beta,\tau) = \int_{-\infty}^{\infty} \frac{(\exp(-\exp(x'\beta \pm g(\varepsilon)))) \exp(y(x'\beta \pm g(\varepsilon)))}{y!} f(\varepsilon;\tau) d\varepsilon \qquad (3.4)$$

for some parameter vector $\tau$. The half normal distribution is, in fact, the distribution of the absolute value of a normally distributed random variable. Therefore, it follows that our model may be written as

$$\mathbb{P}(y|x;\theta) = \int_{-\infty}^{\infty} \frac{(\exp(-\exp(x'\beta \pm |\varepsilon|))) \exp(y(x'\beta \pm |\varepsilon|))}{y!} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}) d\varepsilon \quad (3.5)$$

where the integral is now over the whole $\mathbb{R}$. A change of variable identical to that yielding equation (3.2) leads to the approximate conditional distribution of $y$, from which the log-likelihood may be obtained. The ensuing log-likelihood is

$$\mathcal{L} \propto \sum_{i=1}^{n} \log \sum_{j=1}^{q} \mathbb{P}(y_i|x_i;\theta;\xi_j) \qquad (3.6)$$

where the nodes and weights in the inner sum depend on which of the two methods has been chosen. When the model is an adequate representation of the underlying data generation process, maximizing the log-likelihood over $\theta$ produces consistent, asymptotically normal and efficient estimates of these parameters, and the standard inferential procedures (Score, LR and Wald tests) can be applied in the usual manner (White (1982)).

The moments of the conditional distribution can be obtained now. Given the density function $f(\varepsilon;\sigma_\varepsilon)$, it is not difficult to show (and this is done in Appendix A) that the transformation $\exp(+\varepsilon)$ has density and first two moments:

$$\begin{aligned} f(u) &= f(\exp(\varepsilon)) = \frac{2}{u\sigma_\varepsilon\sqrt{2\pi}} e^{-\frac{(\log(u))^2}{2\sigma_\varepsilon^2}} \mathbb{I}_{[1,\infty)} && (3.7) \\ \mathcal{E}(u) &= e^{\sigma^2/2} \left\{ 1 + \mathrm{Erf}\left(\frac{\sigma_\varepsilon}{\sqrt{2}}\right) \right\} \geq 1 && (3.8) \\ \mathcal{V}(u) &= e^{2\sigma_\varepsilon^2} \left\{ 1 + \mathrm{Erf}\left(\sigma_\varepsilon\sqrt{2}\right) \right\} - \left\{ e^{\sigma_\varepsilon^2/2} \left\{ 1 + \mathrm{Erf}\left(\frac{\sigma_\varepsilon}{\sqrt{2}}\right) \right\} \right\}^2 && (3.9) \end{aligned}$$

where Erf is the error function, defined in the appendix and $\mathbb{I}_.$ is an indicator function. It follows that $f(u)$ is now twice a log-normal distribution with domain on $u \geq 1$, and since the domain of $u$ is restricted to $[0,1]$ it shifts the parameter $\lambda$ downwards, causing the underproduction of counts expected from inefficiency in the production of economic goods. Similarly, for the case of economic goods we have:

$$
\begin{aligned}
f(u) &= f(\exp(-\varepsilon)) = \frac{2}{u\sigma_\varepsilon\sqrt{2\pi}} e^{-\frac{(\log(u))^2}{2\sigma_\varepsilon^2}} \mathbb{I}_{[0,1]} && (3.10)\\
\mathcal{E}(u) &= e^{\sigma_\varepsilon^2/2}\mathrm{Erfc}\left(\frac{\sigma_\varepsilon}{\sqrt{2}}\right) \in [0,1] && (3.11)\\
\mathcal{V}(u) &= e^{2\sigma_\varepsilon^2}\mathrm{Erfc}\left(\sigma_\varepsilon\sqrt{2}\right) - \left\{e^{\sigma_\varepsilon^2/2}\mathrm{Erfc}\left(\frac{\sigma_\varepsilon}{\sqrt{2}}\right)\right\}^2. && (3.12)
\end{aligned}
$$

where now and Erfc is the complementary error function (see Appendix A). Note that the density functions are identical, and they only differ on their domains. Both distributions have well defined mean and variances, from which it follows that, for $u = exp(\varepsilon)$,

$$
\begin{aligned}
\mathcal{E}(y|x) &= \lambda e^{\sigma_\varepsilon^2/2}\left\{1 + \mathrm{Erf}\left(\frac{\sigma_\varepsilon}{\sqrt{2}}\right)\right\} && (3.13)\\
\mathcal{V}(y|x) &= \mathcal{E}(y|x)\left[1 + \lambda W\right] && (3.14)
\end{aligned}
$$

where

$$
W = \frac{e^{2\sigma_\varepsilon^2}\left\{1 + \mathrm{Erf}\left(\sigma_\varepsilon\sqrt{2}\right)\right\} - e^{\sigma_\varepsilon^2/2}\left\{1 + \mathrm{Erf}\left(\frac{\sigma_\varepsilon}{\sqrt{2}}\right)\right\}}{e^{\sigma_\varepsilon^2/2}\left\{1 + \mathrm{Erf}\left(\frac{\sigma_\varepsilon}{\sqrt{2}}\right)\right\}}
$$

and similarly for the case of economic goods. In both cases, the first two moments of the conditional distribution of $y$ are well defined: they exist and their existence does not hinge on the value of the parameters. This seemingly trivial property is, however, not shared by all possible models, as we discuss next.

## 3.1 A Model with a Gamma random effect

The CDSF model just introduced has a number of strengths. First it is a well behaved distribution, with well defined first two moments. Also its flexibility permits various patterns of inefficiency in a sample, thanks to the variance parameter of the mixing distribution. However, in the same fashion as Greene (1990), we wonder whether additional flexibility can be achieved by using a two parameter distribution, such as the Gamma distribution, and whether the resulting conditional distribution for $y$ is a suitable model for CDSF. The answer is only partially yes, because of the asymmetric treatment of economic goods and bads.

Suppose that $\varepsilon$ has gamma distribution with parameters $\alpha > 0$ and $\delta > 0$ (so that $f(\varepsilon) = \frac{\delta^\alpha}{\Gamma(\alpha)}\varepsilon^{\alpha-1}e^{-\delta\varepsilon}$). Then $y$ is conditionally distributed as:

$$
\begin{aligned}
\mathbb{P}(y|x) &= \int_0^\infty \frac{(\exp(-\exp(x'\beta \pm \varepsilon)))\exp(y(x'\beta \pm \varepsilon))}{y!}\frac{\delta^\alpha}{\Gamma(\alpha)}\varepsilon^{\alpha-1}e^{-\delta\varepsilon}d\varepsilon\\
&= \frac{1}{\Gamma(\alpha)}\int_0^\infty \frac{(\exp(-\exp(x'\beta \pm \frac{\epsilon}{\delta})))\exp(y(x'\beta \pm \frac{\epsilon}{\delta}))}{y!}\epsilon^{\alpha-1}e^{-\epsilon}d\epsilon. && (3.15)
\end{aligned}
$$

This distribution does not have closed form, but it could be used in practice by approximating the above integral via Gauss-Laguerre quadrature (see Press et al. (1992))[3]

The problem with this model is that when $y$ is an economic bad, so that $u = \exp(\varepsilon)$, the distribution of $y$ might lack moments. Under the assumed gamma distribution, the density function of the transformation is

$$f(u) = f(\exp(\varepsilon)) = \frac{\delta^\alpha}{\Gamma(\alpha)} \frac{(\log(u))^{\alpha-1}}{u^{\delta+1}} \text{ for } u \in [1, \infty). \tag{3.16}$$

Consider now, for example, the case when we let $\delta = \alpha$. In this case, moments of order $\delta$ or above won't exist, since the integral $\int_1^\infty \log(x)/x dx$ is not convergent. This lack of moments will then carry over to the conditional distribution of $y$. This follows from the well known result that, if $Z$ has a Mixed Poisson distribution with mixing parameter $\theta$, then its moments are

$$\mathcal{E}(Z^r) = \sum_{s=1}^{r} S(r,j)E(\theta^j) \text{ for } j = 1, 2, \ldots \tag{3.17}$$

(see Karlis and Xekalaki (2005)) where $S(.,.)$ are Stirling's numbers of the second kind. An even more illustrative case is that relating the exponential distribution (a type of degenerate Gamma distribution). If $\varepsilon$ has an exponential distribution, then $\exp(\varepsilon)$ has a Pareto distribution on $[1, \infty)$[4]. It can be shown[5] that the expected value and variance of the ensuing Mixed Poisson model for $y$ is

$$\mathcal{E}(y|x) = \lambda \frac{\theta}{\theta - 1} \tag{3.18}$$

$$\mathcal{V}(y|x) = \lambda \frac{\theta}{\theta - 1} \left\{ 1 + \lambda \left\{ \frac{\theta - 1}{\theta - 2} - \frac{\theta}{\theta - 1} \right\} \right\} \tag{3.19}$$

In the above equation, $\theta$ is a parameter inherited from the exponential distribution of $\varepsilon$ and it is originally assumed non-negative. It is clear that the variance exists if an only if $\theta > 2$; otherwise only the first moment is well defined (provided $\theta > 1$). Higher order moments require further restrictions on $\theta$. If $\theta < 1$ then the distribution has no moments.

The above discussion implies that plausible distributional choices for $f(\varepsilon)$ might lead to ill behaved distributions when $y$ is an economic bad -but not when $y$ is an economic good[6]. Because of this, we will say that these distributions do not treat economic goods

[3] We note at this point that the Normal-Gamma stochastic frontier model (Greene (1990), Greene (2003)) was originally estimated via Newton-Cotes quadrature and Simulation methods. However, Judd (1998) reminds us that Gaussian quadrature methods can approximate integrals more accurately than Newton-Cotes methods. Also it transpires that Gaussian quadrature for a single integral can be as accurate as (Pseudo) Monte Carlo integration and certainly much less computer intensive. Therefore, we suggest Gauss-Laguerre quadrature as a third method of estimating Greene's Normal-Gamma SFM.

[4] See, among others, Mood et al. (1974)

[5] A proof of this claim is available from the authors upon request

[6] We could not find instances where the converse was true, however we do not rule out such a possibility, since our analysis in this respect has not been exhaustive.

and bads symmetrically. In practice, these asymmetric models can be implemented provided that $y$ represents an economic good. However, it is implicit in the literature that a proper SFM treats goods and bads symmetrically and, therefore, we are reluctant to include these other models within the class of CDSF discussed in this article.

# 4 Introducing Cross-Sectional Heterogeneity

Until recently, a recurrent argument regarding the weaknesses of (continuous data) parametric stochastic frontier models (whether in their distribution free, as in Schmidt and Sickles (1984), or ML formats) was their inability to separate inefficiency from what is, in fact, unmeasured cross-sectional heterogeneity. In the context of a complete econometric model of a production function, this cross-sectional heterogeneity refers to firm specific characteristics which are unmeasurable and which relate to things other than inefficiency such as the technology in use.

Because ignored heterogeneity is a form of misspecification, the maximum likelihood estimator converges to a pseudo-true parameter value distinct from the one of interest for the researcher and furthermore, the asymptotic variance needs to be adjusted. In our context, this has implications for the interpretation of the estimated coefficients of the production function, but also for the estimates of efficiency, all of which will be biased and inconsistent. The CDSF discussed in the previous section, with its single random effect, is applicable to those data sets where neglected heterogeneity is not likely to be an issue; otherwise, this component should be incorporated into the analysis.

In order to capture its effect, this article assumes that unobserved heterogeneity enters the model additively, so that we introduce a second error component in the mean of the Mixed Poisson distribution associated with $y$. Thus, the conditional mean of the model in equation (2.4) is now:

$$\tilde{\lambda} = \exp(x'\beta \pm |\varepsilon| + \nu) \tag{4.1}$$

where, $\nu \in \mathbb{R}$, $\nu \sim f(\nu;\eta)$ ($\eta$ a vector of real numbers) and $\mathcal{E}(\nu) = 0$. The latter assumption reflects the view that unobserved heterogeneity is likely to accumulate a variety of positive and negative effects, but on average we do not expect its overall effect to shift the production frontier in either direction. When affecting the sample, the model in Section 3 will compound the effect of additive heterogeneity in the random variable capturing inefficiency; this will result in misleading efficiency estimates (see the following section) and it is expected that the structural parameters of the conditional mean will be also biased. We explore this in a simulation experiment in a later section.

There are, of course, many candidates for $f(\nu;\eta)$, but the normal distribution $N(0,\sigma_\nu^2)$ seems a natural choice, and although it won't facilitate a closed form log-likelihood function, we still can approximate the ensuing conditional distribution of $y$ via Gauss-Hermite quadrature, as is shown below. Under the assumption of normality, the conditional distribution of $y$ is

$$\mathbb{P}(y|x;\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Po(y|x,\varepsilon,\nu)\frac{1}{\sigma_\varepsilon\sigma_\nu 2\pi}\exp\left\{-\frac{1}{2}\left(\frac{\varepsilon^2}{\sigma_\varepsilon^2}+\frac{\nu^2}{\sigma_\nu^2}\right)\right\} d\varepsilon d\nu \quad (4.2)$$

$$Po(y|x,\varepsilon,\nu) = \frac{(\exp(-\exp(x'\beta \pm |\varepsilon| + \nu)))\exp(y(x'\beta \pm |\varepsilon| + \nu))}{y!} \quad (4.3)$$

where now $\theta' = (\beta', \sigma_\varepsilon, \sigma_\nu)$. As before, a simple change of variable allows us to express the above function in terms of an integral of the form $\int_{\mathbb{R}}\int_{\mathbb{R}} h(x_1, x_2)\exp(-x_1^2 - x_2^2)$, which can be approximated by Cartesian products of Gauss-Hermite formulas (see Stroud (1971)). The log-likelihood function that must be maximized is:

$$\mathcal{L} = \sum_{i=1}^{n}\log(L_i) = \sum_{i=1}^{n}\log\sum_{j=1}^{q}\sum_{k=1}^{q}\frac{1}{\pi}Po(y_i|x_i,\xi_j,\xi_k)w_j w_k \quad (4.4)$$

with

$$Po(y_i|x_i,\xi_j,\xi_k) = \frac{(\exp(-\exp(x_i'\beta \pm \sqrt{2}\sigma_\varepsilon|\xi_j| + \sqrt{2}\sigma_\nu\xi_k)))\exp(y_i(x_i'\beta \pm \sqrt{2}\sigma_\varepsilon|\xi_j| + \sqrt{2}\sigma_\nu\xi_k))}{y_i!}$$

and where $\xi_j$ and $w_j$ are the nodes and weights from standard Gauss-Hermite quadrature. The analytical derivatives of $\mathcal{L}$ are given by

$$\frac{\partial\mathcal{L}}{\partial\theta} = \frac{1}{\pi}\sum_{i=1}^{n}\frac{1}{L_i}\sum_{j=1}^{q}\sum_{k=1}^{q}Po(y_i|x_i,\xi_j,\xi_k)(y_i - \tilde{\lambda}_{i,j,k})\left\{\begin{array}{c} x_i \\ \pm\sqrt{2}|\xi_j| \\ \sqrt{2}\xi_k \end{array}\right\} \quad (4.5)$$

where $L_i$ was defined in equation (4.4). The gradient of model (3.6) follows from the above expression in a straightforward manner.

The reader has probably noted the similarity between this and the ALS models in the sense that the random effects have identical distributions (convoluting a normal and half normal random variables). For this reason, and given the importance of the contribution found in Aigner et al. (1977), we judge it appropriate to name this a Poisson-Aigner-Lovell-Schmidt model despite of its colorful acronym: P.A.L.S..

In our experience with this model, the use of analytical derivatives with maximization routines such as the BFGS has substantially improved the speed of the algorithms and, of course, this approach has also facilitated the computation of standard errors via the Outer Product of Gradients estimator (also called BHHH estimator) -as opposed to using the minus inverse of the Hessian Matrix. Nonetheless, it is convenient to remark that, should we wish to construct a likelihood based test of some sort, evidence exists regarding the diminished small sample properties of OPG-based tests of hypothesis (see, for example, Orme (1990)).

Because this model nests the PHN, Poisson-Log-Normal (PLN) and Poisson distributions, test of hypothesis can be easily performed by means of likelihood ratios. For example, to test the null hypothesis of no-heterogeneity ($H_0 : \sigma_\nu^2 = 0$), one only needs to estimate both PHN and PALS models and calculate twice the difference between the log-likelihoods, which then can be compared with the quantiles of a $\chi_1^2$ distribution. The null hypothesis of no inefficiency ($H_0 : \sigma_\varepsilon^2 = 0$) would be based on the log-likelihoods of the PHN and PLN model, and so on. However, in contrast with Aigner et al. (1977), the ratio of the standard errors of each random effect provides information regarding the relative importance of inefficiency and heterogeneity in explaining the overdispersion in the sample (as opposed to measuring the relative importance of inefficiency in the observed total variation in the sample). Thus,

$$\eta = \frac{\sigma_\varepsilon}{\sigma_\nu} \tag{4.6}$$

then $\eta > 1$ would point to inefficiency as the main cause of the overdispersion in the sample, whereas $0 < \eta < 1$ points out the predominance of unobserved heterogeneity.

Finally, it is intuitively clear that adding a second zero-mean random effect in the model leaves the conditional mean of $y$ in equations (3.11) and (3.13) unaffected; on the other hand, although the conditional variance is affected by the inclusion, it is still well defined for both economic goods and bads, and thus this model also treats all commodities symmetrically.

# 5 Estimation of Inefficiency

Although the parameters of the production frontier are of interest in themselves, the ultimate goal of our analysis is to obtain approximate efficiency scores for each individual in the sample. To attain this goal, we rely on the method found in Jondrow et al. (1982), by which individual efficiency scores are provided by the posterior expectation $\mathcal{E}(u|y,x)$. From Bayes' theorem we know that

$$f(u|x,y) = \frac{\mathbb{P}(y|x,u)f(u)}{\mathbb{P}(y|x)} \tag{5.1}$$

so that the posterior expected value of $u$ is simply $\int uf(u|x,y)du$. As expected, none of the above models has a closed form for the conditional distribution of $u$ given $x, y$; however we may still approximate the relevant integrals by their quadrature approximations. Thus, the approximate $\mathcal{E}(y|x)$ for the Mixed Poisson model with Half Normal random effect is

$$\hat{u}_i \quad = \quad \mathcal{E}(u_i|x_i, y_i) \approx \frac{\sum_{j=1}^{J} e^{\pm\xi_j\sigma\sqrt{2}} Po(y_i|x_i;\xi_j)w_j}{\sum_{j=1}^{J} Po(y_i|x_i;\xi_j)w_j} \tag{5.2}$$

and similarly for the double random effect model. The above quantities are unfeasible, since they depend on the unknown parameters of the model, however feasible estimates would follow by using the maximum likelihood estimates of $\beta$ and $\theta$.

At this junction, the remarks by Wang and Schmidt (2009) seem pertinent. The distributions of $u$ and $\hat{u}$ are not the same, as can be seen by noting that $\mathcal{V}(u) =$

$\mathcal{V}(\mathcal{E}(u|x,y)) + \mathcal{E}(\mathcal{V}(u|x,y))$, (and hence, $\hat{u}$ has smaller variance) but $\mathcal{E}(u) = \mathcal{E}(u|x) = \mathcal{E}(\mathcal{E}(u|x,y))$. Furthermore, $\hat{u}$ is a shrinkage[7] of $u$ toward its mean. As a result, the lower and upper tails of the distribution of $u$ will be miss-reported, so that $\hat{u}$ penalizes outstanding firms and rewards the least efficient individuals -although the average efficiency in the sample is correctly approximated. Quoting Wang and Schmidt (2009) this *...does not mean that there is anything wrong with the estimator* since ... *it is unbiased in the unconditional sense* $E(\hat{u} - u) = 0$. But this characteristic must be taken into account in applications.

## 6 Evidence from a Monte Carlo Experiment

In order to evaluate the small sample merits of the above quadrature procedures and to gain an indication of the sort of biases that neglected unobserved heterogeneity might induce, we performed a limited Monte Carlo simulation. In our experiment, we generated data from two models:

$$DGP1: \quad \mathbb{P}_1(y|x,\varepsilon) = \text{Poisson}(\exp(x'\beta - |\varepsilon|)) \tag{6.1}$$

$$DGP2: \quad \mathbb{P}_2(y|x,\varepsilon,\nu) = \text{Poisson}(\exp(x'\beta - |\varepsilon| + \nu)). \tag{6.2}$$

The covariate vector was $x' = (1, x_1, x_2)$, where $x_1 \sim \text{Uniform}[0,1]$ and $x_2 = \rho x_1 + \sqrt{1-\rho^2}V_i$, where $V \sim \text{Uniform}[0,1]$, $\rho = 0.5$, and the vector of parameters $\beta$ was set at $(0.5; 0.5; 0.5)'$. The distributions of $\varepsilon$ and $\nu$ were $N(0,1)$. Although we focus our attention on cases when $y$ is an economic good, similar results were obtained for the case of economic bads.

Once simulated data sets were available, CDSF models were specified in which the structural part of the conditional mean $\tilde{\lambda}$ was correctly specified, so that $\lambda = \exp(x'\beta)$[8]. We used samples of 500, 1000 and 2000 observations, and each model was estimated 200 times. In each model we used 20 points of quadrature, although we found results to be rather robust to the number of quadrature points, and indeed 10 points of quadrature could have been safely used. Simulations were run in Ox v.4. with OxEdit as interface, and maximization was based on `MaxSQP` -a maximizer allowing linear restrictions; in particular $\sigma_\varepsilon$ and $\sigma_\nu$ were restricted to $\mathbb{R}^+$, as is natural. The quadrature algorithm in our simulations was an straightforward adaptation of that in Press et al. (1992). The mean and standard errors of the estimated parameters for each model are given in the Tables 1 and 2.

Table 1 collects the results of estimating data generated by a Poisson Half Normal distribution, so that heterogeneity was not an issue in estimation. The average value of the estimated parameters is very close to the true parameter value, even for samples of just 500 observations. The precision with which the parameters are estimated increases with the sample size, however we do not observe major variations in the average estimated value of the parameters. It is also noticeable that because of the utilization of

[7]See, for instance, Gourieroux and Monfort (1995))

[8]In this setting, the Poisson Half Normal (PHN hereafter) model is only appropriate with $\mathbb{P}_1(.)$, and it is misspecified if data are produced with $\mathbb{P}_2(.)$

the `MaxSQP` routine in order to rule out negative variances, the estimated value of $\sigma_\nu$ returned by the Poisson Normal Half Normal model was systematically zero -which is, on the other hand, a desirable side effect. As a result, the estimates of the remaining parameters are identical to those provided by the simpler Poisson Half Normal model, since PALS reduces to PHN whenever $\sigma_\nu = 0$.

Table 2 provides information regarding the biases induced by unobserved heterogeneity. The upper half of the table collects the average estimated values and standard errors produced by the Poisson Half Normal model. There is a positive bias, which does not decrease with the sample size. Thus, the estimated average value of $\beta_j$ $(j = 1, 2)$ is about 40% above its true value. The situation is worse with the estimates of the intercept and $\sigma_\varepsilon$. The former exhibits a bias of about +80% while the latter is biased by a value beyond 100%. Misspecification also affects the standar errors, which are substantially inflated with respect to those values displayed in Table 1. In dealing with contaminated data, the single random effect model allocates the excess overdispersion to $\varepsilon$ -which was expected. Even worse, it places the production function at a higher level and exaggerates the elasticities of each covariate.

The second half of Table 2 presents the results obtained by the correctly specified PHN. As expected, the average estimate is concentrated about the true parameter value, although the additional integral in $\mathbb{P}(y|x)$ seems to have a cost in terms of larger standard errors than in the PHN model-but without risking the significance level of the estimated values. Accuracy improves with the sample size, in terms of a smaller standard error and increasing proximity to the true parameter values.

In summary, this simulation suggests that when unobserved cross-sectional heterogeneity is not a modeling issue, either model is capable of producing very similar and efficient results. Furthermore the choice of one or the other can be based on considerations such as computer time, since the two random effects model is much more computer intensive than the single random effect model as it requires $q^2$ points of quadrature to evaluate the double integral (as opposed to just $q$ points required by PHN). On the other hand, if there is a suspicion that unobserved heterogeneity might be an issue, then both models should be computed, and choice of one model or the other should be based on some inferential procedure. Since the PHN model is nested within the PALS model, the likelihood ratio test seems an optimal procedure in order to test the suitability of the former.

# 7 Application to the Relationship Between Patents and R&D at Firm Level

In this section we study the relationship between patents awarded to a firm in a given year and investment in R&D by estimating the production function of patents and, then, we estimate the population's average efficiency level in the production of patents.

The relationship between patents and research and development has been studied at length by a number of authors. Key references are Pakes and Griliches (1980), Hausman et al. (1984), Hall et al. (1986), Pakes (1986), Griliches (1990) and Wang et al. (1998)

among others. The general view, which we adopt here, is that annual expenditure in R&D is a type of investment, and therefore, it enriches the stock of knowledge of a given firm. This knowledge materializes into new technologies, some of which are registered and, subsequently, a patent is awarded to the firm in order to protect the firm's property rights on the invention . Therefore, patents represent the value of the underlying stock of knowledge. The authors mentioned above have acknowledged that patents are not the only output of R&D, and their economic value can be under question, as a large proportion of granted patents are of little economic value or depreciate too quickly (see Pakes (1986)). However, as has been pointed out, firms are awarded patents and this is the product of research and development undertaken within the firm, so that patents measure the levels of research activities -*...even though the information conveyed by an individual patent may be very small* (Hall et al. (1986)).

In the past, a research interest has been the dynamics of R&D expenditure: knowledge stock depreciates over time so contributions of older investment in R&D become less valuable as time goes by. This article has little to say in this respect, and this section looks only the contemporary relationship between investment in R&D and patents awarded. Our focus is the strength of the relationship between R&D and patents, the elasticity of patents with respect to R&D investment and the average effectiveness of the firms in the sample in transforming R&D investment into patents.

Unlike previous research, where the production function was approximated by the expected value of the number of patents conditional on some function of R&D, in this article we use the CDSF models described above to approximate the production function as the curve enveloping the data *from above*. This curve agrees better with the idea of the production function as the frontier of production possibilities, capturing maximum output given the levels of input. Once the production function is estimated, the efficiency measure discussed in section 5 may be used to evaluate how well firms transform R&D into patents.

We use 70 pharmaceutical firms from the 1976 wave of the National Bureau of Economic Research R&D Masterfile (Hall et al. (1986)) in our analysis. By restricting our attention to a single industry, we are reducing the chances of neglected inter-industry heterogeneity (this is of particular importance for obtaining unbiased estimates of the average efficiency in the sample, since any unobserved heterogeneity will be compounded in the mixing variable of the underlying conditional Poisson model, thus distorting our estimates of what proportion of the observed variation in the sample is due to inefficiency). However it is unclear a priori whether this is sufficient to eliminate other forms of neglected heterogeneity. This same dataset was used by Wang et al. (1998)[9] and it includes the count of patents for each firm, as well as the R&D expenditure for that year, and the value of sales in each firm. The summary statistics are in Table 3.

The model we fit assumes that each observation is independently distributed, with a conditional distribution function given by the general model at the end of Section 2.

[9]We thank professors Peiming Wang, Iain Cockburn and Martin Puterman for kindly providing us with their data set

The structural part of the conditional mean is

$$\lambda_i = \exp(\beta_0 + \beta_1 \log(\text{R\&D}_i) + \beta_2 \log(\text{R\&D}_i)^2 + \beta_3 \log(\text{Sales}_i) \tag{7.1}$$

where we have included a quadratic term in R&D in order to take into account further non-linearities. The results produced by our models are given in Table 4. The table reveals significant differences between the Poisson Half Normal and Normal Half Normal models, which are supported by a Likelihood Ratio test exceeding the quantile corresponding to 1% critical region of the $\chi^2_1$ distribution ($H_0 : \sigma_\nu = 0$). After discounting unobserved heterogeneity, the PALS model increases the absolute magnitude of $\sigma_\varepsilon$, while it places the intercept of the frontier at a lower level. In relative terms, PALS suggests that the variation in the sample due to inefficiency is larger than that due to heterogeneity by a factor of 1.17. Thus, despite the fact that the sample contains a relatively homogeneous population of just pharmaceutical firms, heterogeneity is still a structural feature in the population, and we conclude that the results of the PHN model are likely to be biased. The quality of fit is even more apparent in Figures 1 and 2, where the estimated frontier is drawn against the observed data. It is visually apparent that the PHN model is unable to produce a sufficiently tight fit at the upper tail of the distribution of log(R&D), unlike the PALS model which envelopes the data from above in a manner which truly resembles the frontier which was sought.

The estimated coefficients of log(R&D) and log(R&D)$^2$ can be used in order to calculate the ratio:

$$\frac{\partial \lambda_i}{\partial \log(\text{R\&D})} \frac{1}{\lambda_i} = \beta_1 + 2\beta_1 \log(\text{R\&D}) \tag{7.2}$$

which measures the percentage variation in the frontier as a fraction of the variation in the logarithm of R&D (it is, therefore, a pseudo elasticity). The average value of this ratio provided by PHN model is 1.1341 (s.e.: 0.49310) as opposed to the average value of 1.0452 (s.e.: 0.18203) reported by PALS. Given the potential biases inherent in the PHN we focus our attention on the latter value, which is moderately optimistic regarding the multiplicative effect of R&D. A simple Wald test does not reject the hypothesis of the above ratio being equal to 1, and therefore, there is some evidence that we might be facing a situation of constant returns to R&D.

Table 5 summarizes the estimated posterior distribution of the half normal random effect, which is the basis for our analysis of efficiency in the sample. Average efficiency is at levels of 0.59% (PHN) and 0.64% (PALS). The corresponding median values are 0.62 and 0.64. The distribution generated by the PHN model is skewed, suggesting that that approximately 10% of the sample is unable to achieve efficiency levels above 34% while at the other end of the spectrum, 25% of the firms in the sample may attain excellent levels of efficiency, ranging above 70%. However, previous considerations, reinforced by Figure 1, invite us to reconsider the adequacy of these estimates. In the figure, the output of those firms with largest investment in R&D lie at further distance from the estimated frontier than those with lowest R&D expenditure. Since, ultimately, our estimates of efficiency rely on the vertical distance between the frontier and the observed value of output, we conclude that the PHN, (which neglects potential cross-sectional heterogeneity) is, in this case, penalizing the firms with largest R&D investment. On the contrary, the PALS provides more balanced evidence. The estimated distribution of efficiency is fairly symmetric about the mean, and in accordance with the evidence provided by this model, only 10% of the firms in the sample would attain efficiency levels below 52%. On the other end, the $99^{th}$ quantile of the distribution is located at an efficiency level of 76%. Figure 2 exhibits a much tighter fit in the upper tail of the sample, and this provides further evidence on the bias of the PHN model caused by ignored heterogeneity.

Looking at the estimated distribution of efficiency, the PALS model may seem rather pessimistic regarding the capability of these pharmaceutical firms in order to produce patents. However, it is necessary to remark on two of the points discussed above. Firstly, in the way it has been defined, the estimator of efficiency is a shrinkage toward the mean of $u$ (Wang and Schmidt (2009)), and we would thus expect the upper and lower tails of the estimated distribution to be more concentrated about the mean (of $u$) than the actual distribution of inefficiency. However, this would not explain the low estimated average efficiency. Perhaps, a more important point relevant to interpreting this result is the fact that patents are not the only output of R&D, and that their economic value can be under question. Since not all R&D needs to materialize in patents, it is normal to expect the estimator in Section 5 to return a low average value of efficiency.

To summarize we find some evidence suggesting that there are contemporary constant returns to scale of R&D in the production of patents, however, this view is in contrast with the typically low achievement in terms of transforming R&D into patents. This latter point need not be surprising, since patents are not the only outcome from R&D and in some occasions the economic value of patents may be negligible (eliminating the incentive to apply for a patent).

# 8 Conclusion

Existing contributions to the area of stochastic frontier analysis have been devised with a continuous output variable in mind. However, one often encounters situations when the output variable of interest is a count, so that its probability distribution function is discrete. While econometricians have been proposing and improving for decades estimators for the analysis of mean regression of count data , no attempt has been made to propose a Stochastic Frontier method for a discrete dependent variables. This article completes the existing literature by putting forward a proper Count Data Stochastic Frontier model.

Our discussion has been focused on producing a model that can treat economic goods and bads symmetrically, in the sense that the ensuing conditional distribution of output has at least well defined low order moments, and that the existence of these moments is not subject to the value of the parameters in the model. Whether the final distribution has lower order moments depends on what distribution has been assumed for the the random variable that caputures inefficiency in the sample. We have shown that while the half-normal distribution does result in a proper CDSF model, other more adaptable choices (for instance, the gamma distribution) lead to ill behaved distributions for the output variable.

We have emphasized the role played by unobserved cross-sectional heterogeneity, and we have shown that when this part of the model is neglected, estimation can be severely biased. We have shown how to make our models robust to unobserved cross-sectional heterogeneity in such a fashion that inefficiency and heterogeneity are treated as separate sources of variation. This is a rather unique feature in the literature of Stochastic Frontiers, and it is only shared by a handful of model, among which Greene (2005) is of particular interest.

We have illustrated the scope of our models by applying them to a substantive problem: the study of the relationship between R&D expenditure and number of patents awarded to a firm, as an measure of the repercussion of that type on investment on the accumulation of knowledge in the firm. Our approach is different to that in key early contributions, not only because of the use of a CDSF model, but because of the emphasis on estimation of a frontier of production of patents as opposed to a mean regression curve. Unlike other contributions, we also explore the extent to which R&D is transformed into patents, via the efficiency measure introduced in this article.

While developing our methods we assumed that inefficiency and heterogeneity are independently distributed. This might seem a rather restrictive assumption and, indeed, it will be so in certain occasions. Then, Maximum Likelihood estimation of our models would return inconsistent estimates. A similar consideration led Smith (2007) to relax an equivalent assumption in the case of continuously distributed output. He relied on a copula method in order to capture the magnitude of the correlation between inefficiency and the regression error in the model. In our context, this approach seems to be equally attractive and feasible, but if this line of work is followed care should be taken that the resulting model treats economic goods and bads symmetrically, in the sense we have emphasized here. Otherwise, one may exploit the assumed normality of the additive random effects, and modify the models discussed in this article by introducing the pertinent bivariate normal distribution. We presume this would still result in symmetric models, but feeling that this question it is beyond the scope of this article, we leave that for future research.

## References


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

Battese, G. and T. Coelli (1995). A model for technical inefficiency effects in stochastic frontier production frontier for panel data. *Empirical Economics 20*, 325–332.

Butler, J. and R. Moffitt (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica 50*, 761–764.

Cameron, C. and P. Trivedi (1998). *Regression analysis of count data*. Cambridge University Press.

Cornwell, C., P. Schmidt, and R. Sickles (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics 46*, 185–200.

Fader, P. and B. Hardie (2000). A note on modelling underreported poisson counts. *Journal of Applied Statistics 8*, 953–964.

Fé-Rodríguez, E. (2007). Exploring a stochastic frontier model when the dependent variable is a count. *The School of Economics Discussion Paper Series, University of Manchester 0725.*

Fé-Rodríguez, E. (2008). On the production of economic bads and the estimation of production efficiency when the dependent variable is a count. *The School of Economics Discussion Paper Series, University of Manchester 0812.*

Gentle, J. E. (2003). *Random Number Generation and Monte Carlo Methods, 2nd Edition*. Springer.

Geweke, J. (1995). Monte carlo simulation and numerical integration. *Federal Research Bank of Minneapolis Research Report 192*.

Gourieroux, C. and A. Monfort (1995). *Statistics and Econometric Models*. Cambridge University Press.

Gozalo, P. and O. Linton (2000). Local nonlinear least squares: Using parametric information in nonparametric regression. *Journal of Econometrics 99*, 63–106.

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

Greene, W. (2003). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis 19*, 179–190.

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

Griliches, Z. (1990). Patent statistics ad economic indicators: A survey. *Journal of Economic Literature 28*, 1661–1707.

Hall, B., C. Cummins, E. Laderman, and J. Murphy (1986). The r&d master file documentation. *Technical Working Paper, NBER 72.*

Hall, B., Z. Griliches, and J. A. Hausman (1986). Patents and r&d: Is there a lag? *International Economic Review 27*, 265–283.

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

Hofler, R. and D. Scrogin (2008). A count data stochastic frontier. *Discussion Paper, University of Central Florida.*

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

Judd, K. L. (1998). *Numerical Methods in Economics.* The MIT Press.

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

Kumbhakar, S., B. U. Park, L. Simar, and E. Tsionas (2007). Nonparametric stochastic frontier: A local maximum likelihood approach. *Journal of Econometrics 137*, 1–27.

Kumbhakar, S. C. and K. Lovell (2003). *Stochastic Frontier Analysis.* Cambridge University Press.

Martins-Filho, C. and F. Yao (2007). Nonparametric frontier estimation via local linear regression. *Journa of Econometrics 141*, 283–319.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995–1026.

Mood, A., F. Graybill, and D. Boes (1974). *Introduction to the Theory of Statistics.* McGraw Hill.

Orme, C. (1990). The small sample performance of the information matrix test. *Journal of Econometrics 46*, 309–331.

Pakes, A. (1986). Patents as options: Some estimates of the value of holding european patent stocks. *Econometrica 54*, 755–784.

Pakes, A. and Z. Griliches (1980). Patents and r&d at the firm level: A first report. *Economic Letters 5*, 377–381.

Park, B., R. Sickles, and L. Simar (2007). Semiparametric efficient estimation of dynamic panel data models. *Journa of Econometrics 136*, 281–301.

Park, B. U., R. Sickles, and L. Simar (1998). Stochastic panel frontiers: a semiparametric approach. *Journa of Econometrics 84*, 273–301.

Press, W., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C (The art of Scientific Computing), Second Edition.* Cambridge University Press.

Schmidt, P. and R. Sickles (1984). Production frontiers and panel data. *Journal of Business and Economic Statistics 2*, 367–374.

Sickles, R. (2005). Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. *Journal of Econometrics 126*, 305–334.

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

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

Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals.* Prentice Hall.

Wang, P., I. M. Cockburn, and M. L. Puterman (1998). Analysis of patent data: A mixed-poisson-regression-model approach. *Journal of Business and Economic Statistics 16*, 27–41.

Wang, W. and P. Schmidt (2009). On the distribution of estimated technical efficiency in stochastic frontier models. *Journal of Econometrics 148*, 36–45.

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

Winkelmann, R. (2008). *Econometric Analysis of Count Data (Fifth Edition).* Springer.

## Appendix A: Moments of the density function $f(\exp(\pm\varepsilon))$

We now show how to calculate the first and second moments of the transformation $u = f(\exp(\pm\varepsilon))$, where $\varepsilon$ follows a half normal distribution. Consider first the case $f(\exp(-\varepsilon))$. Then,

$$
\begin{aligned}
\mathcal{E}(u) &= \frac{2}{\sigma\sqrt{2\pi}}\int_0^1 e^{-\log^2(u)/2\sigma^2}du = \frac{2}{\sigma\sqrt{2\pi}}\int_{-\infty}^0 e^{-(\frac{t^2}{2\sigma^2}-t)}dt = e^{\sigma^2/2}\frac{2}{\sqrt{\pi}}\int_{-\infty}^{-\frac{\sigma}{\sqrt{2}}} e^{-s^2}ds \\
&= e^{\sigma^2/2}\mathrm{Erfc}\left(\frac{\sigma}{\sqrt{2}}\right) \in [0,1] \qquad \text{(A-1)}
\end{aligned}
$$

where Erfc(.) is the complementary error function and we used the changes of variable $\log(x) = t$, $s = \frac{t}{\sqrt{2}\sigma} - \frac{\sigma}{\sqrt{2}}$ and the fact that $\frac{t^2}{2\sigma^2} - t = \left(\frac{t}{\sigma\sqrt{2}} - \frac{\sigma\sqrt{2}}{2}\right)^2 - \frac{\sigma^2}{2}$. Similar steps show that

$$
\begin{aligned}
\mathcal{E}(u^2) &= e^{2\sigma^2}\mathrm{Erfc}\left(\sigma\sqrt{2}\right) \qquad \text{(A-2)} \\
\mathcal{V}(u) &= e^{2\sigma^2}\mathrm{Erfc}\left(\sigma\sqrt{2}\right) - \left\{e^{\sigma^2/2}\mathrm{Erfc}\left(\frac{\sigma}{\sqrt{2}}\right)\right\}^2 \qquad \text{(A-3)}
\end{aligned}
$$

For the case $u = \exp(\varepsilon)$ the method is identical, but only the range of integration changes to $[1,\infty)$. Thus,

$$
\begin{aligned}
\mathcal{E}(u) &= \frac{\sqrt{2}}{\sigma\sqrt{\pi}}\int_1^\infty e^{-\log^2(u)/2\sigma^2}du = e^{\sigma^2/2}\frac{2}{\pi}\int_{-\frac{\sigma}{\sqrt{2}}}^\infty e^{-s^2}ds \\
&= e^{\sigma^2/2}\left\{1+\mathrm{Erf}\left(\frac{\sigma}{\sqrt{2}}\right)\right\} \geq 1 \qquad \text{(A-4)} \\
\mathcal{E}(u^2) &= e^{2\sigma^2}\left\{1+\mathrm{Erf}\left(\sigma\sqrt{2}\right)\right\} \qquad \text{(A-5)}
\end{aligned}
$$

**DGP 1**

| | Poisson-Half Normal | | | | | |
|---|---|---|---|---|---|---|
| | $N = 500$ | | $N = 1000$ | | $N = 2000$ | |
| | **Mean** | **S.d** | **Mean** | **S.d** | **Mean** | **S.d** |
| $\sigma_\varepsilon$ | 1.0867 | *0.0249* | 1.0872 | *0.0118* | 1.0790 | *0.0054* |
| $\beta_0$ | 0.5827 | *0.0205* | 0.5790 | *0.0102* | 0.5785 | *0.0060* |
| $\beta_1$ | 0.5006 | *0.0296* | 0.5099 | *0.0145* | 0.5008 | *0.0082* |
| $\beta_2$ | 0.4965 | *0.0248* | 0.4937 | *0.0146* | 0.4946 | *0.0078* |

| | Poisson-Normal-Half Normal | | | | | |
|---|---|---|---|---|---|---|
| | $N = 500$ | | $N = 1000$ | | $N = 2000$ | |
| | **Mean** | **S.d** | **Mean** | **S.d** | **Mean** | **S.d** |
| $\sigma_\varepsilon$ | 1.0867 | *0.0249* | 1.087 | *0.0118* | 1.0769 | *0.0057* |
| $\sigma_\nu$ | 0 | 0 | 0 | 0 | 0.0036 | *0.0007* |
| $\beta_0$ | 0.5827 | *0.0205* | 0.5798 | *0.0102* | 0.5768 | *0.0062* |
| $\beta_1$ | 0.5005 | *0.0296* | 0.5099 | *0.0145* | 0.5008 | *0.0082* |
| $\beta_2$ | 0.4964 | *0.0248* | 0.4937 | *0.0146* | 0.4946 | *0.0078* |

Table 1: Monte Carlo Simulation with data generated from DGP 1 (no unobserved cross-sectional heterogeneity)

**DGP 2**

| | Poisson-Half Normal | | | | | |
|---|---|---|---|---|---|---|
| | $N = 500$ | | $N = 1000$ | | $N = 2000$ | |
| | **Mean** | **S.d** | **Mean** | **S.d** | **Mean** | **S.d** |
| $\sigma_\varepsilon$ | 2.4760 | *0.0234* | 2.4471 | *0.0106* | 2.4191 | *0.0035* |
| $\beta_0$ | 1.8974 | *0.2368* | 1.8828 | *0.1196* | 1.843 | *0.0458* |
| $\beta_1$ | 0.7183 | *0.6797* | 0.6894 | *0.3246* | 0.7504 | *0.1064* |
| $\beta_2$ | 0.7072 | *0.6789* | 0.7158 | *0.3264* | 0.7010 | *0.1279* |

| | Poisson-Normal-Half Normal | | | | | |
|---|---|---|---|---|---|---|
| | $N = 500$ | | $N = 1000$ | | $N = 2000$ | |
| | **Mean** | **S.d** | **Mean** | **S.d** | **Mean** | **S.d** |
| $\sigma_\varepsilon$ | 0.7718 | *0.3759* | 0.9698 | *0.2819* | 0.8798 | *0.2338* |
| $\sigma_\nu$ | 1.0289 | *0.0154* | 0.9984 | *0.0110* | 1.0167 | *0.0087* |
| $\beta_0$ | 0.3125 | *0.2520* | 0.4851 | *0.1718* | 0.4265 | *0.1436* |
| $\beta_1$ | 0.5227 | *0.0711* | 0.4879 | *0.0346* | 0.5113 | *0.0201* |
| $\beta_2$ | 0.4921 | *0.0783* | 0.5047 | *0.0380* | 0.4865 | *0.0223* |

Table 2: Monte Carlo Simulation with data generated from DGP 2 (Data generated with unobserved cross-sectional heterogeneity so that PHN is misspecified

**Patents Data Set (Wang et al. (1998));** $N = 70$

| | Variable | | |
|---|---|---|---|
| | Log(R&D) | Log(R&D)$^2$ | Log(Sales) |
| Mean | 1.3119 | 6.0961 | 4.7089 |
| S.D. | 2.0916 | 6.4831 | 2.0361 |
| Max | 4.9152 | 24.160 | 7.8330 |
| Min. | -2.9565 | 0.0017526 | 0.15272 |
| Quantile 10 | -1.7880 | 0.16201 | 1.5864 |
| Quantile 25 | -0.17530 | 0.87232 | 3.6753 |
| Quantile 50 | 1.3106 | 3.3642 | 4.8651 |
| Quantile 75 | 2.9811 | 8.9356 | 6.5212 |
| Quantile 90 | 4.0113 | 16.091 | 7.2076 |

Table 3: Descriptive Statistics

**Estimated Frontier**

*Patents Data Set (Wang et al. (1998)); $N = 70$*

| | Model | | | |
|---|---|---|---|---|
| | Poisson HN | | Poisson ALS | |
| | **Estimate** | **S.d** | **Estimate** | **S.d** |
| $\sigma_\varepsilon$ | 0.55069 | *0.049789* | 0.61073 | *0.14799* |
| $\sigma_\nu$ | - | - | 0.52349 | *0.10629* |
| Intercept | 1.8769 | *0.15547* | 1.8068 | *0.23520* |
| log R&D | 0.82485 | *0.080673* | 0.93102 | *0.10104* |
| log R&D$^2$ | 0.11788 | *0.017096* | 0.043514 | *0.017577* |
| log sales | -0.26867 | *0.035983* | -0.23931 | *0.051626* |
| Log-likelihood | -243.16 | - | -203.424 | - |
| Mean Efficiency | 0.59025 | - | 0.64553 | - |

Table 4: Estimated CDSF for the Wang-Cockburn-Puterman data set

**Estimated Distribution of Efficiency**

*Patents Data Set (Wang et al. (1998)); $N = 70$*

| | Model | |
|---|---|---|
| Quantile | Poisson HN | Poisson ALS |
| 1% | 0.17894 | 0.44513 |
| 10% | 0.34138 | 0.52682 |
| 25% | 0.48674 | 0.60806 |
| 50% | 0.62999 | 0.64971 |
| 75% | 0.71755 | 0.70467 |
| 90% | 0.78075 | 0.73501 |
| 99% | 0.82606 | 0.76544 |
| Mean Efficiency | 0.59025 | 0.64553 |

Table 5: Estimated distribution of the average efficiency score for the Wang-Cockburn-Puterman data set

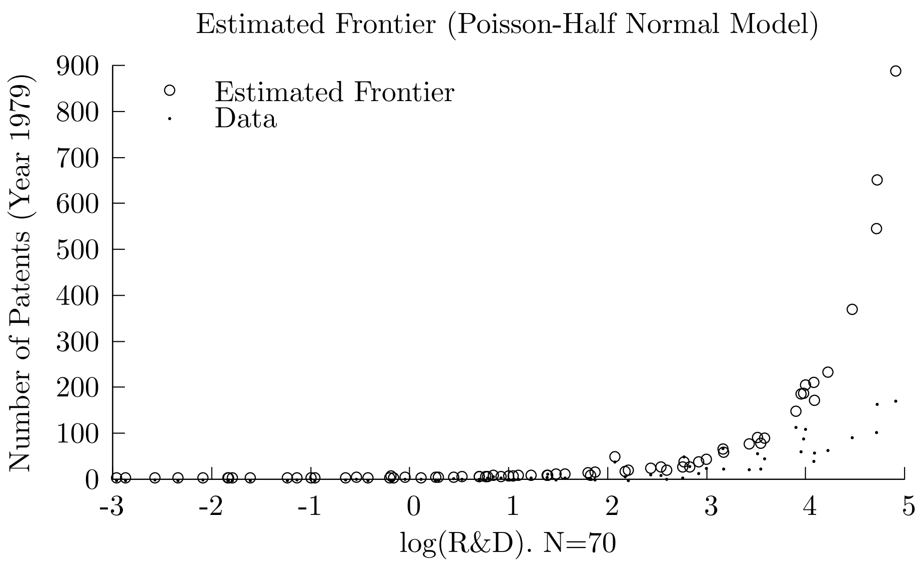


Figure 1: Estimated Frontier (PHN Model)

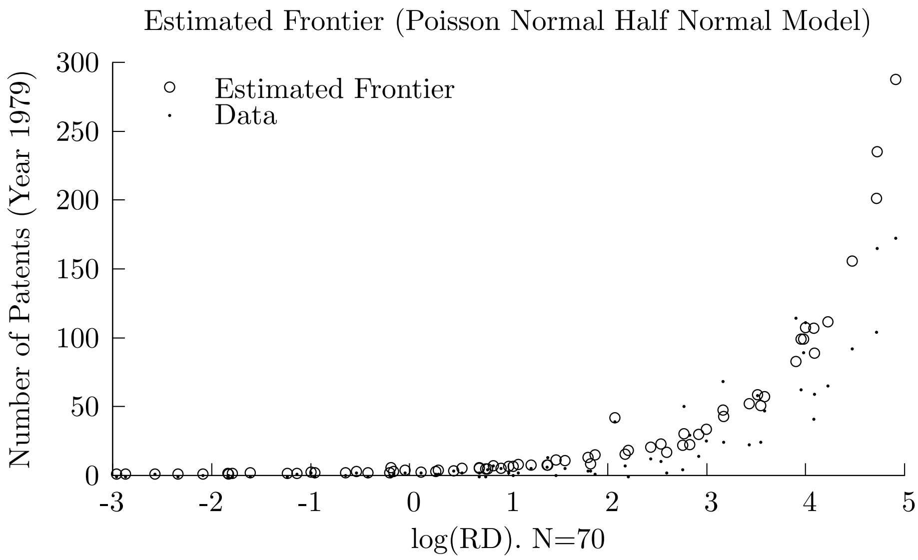


Figure 2: Estimated Frontier (PALS Model)