INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0525

A FULLY NONPARAMETRIC STOCHASTIC FRONTIER MODEL FOR PANEL DATA

D.J. HENDERSON and L. SIMAR

http://www.stat.ucl.ac.be

A Fully Nonparametric Stochastic Frontier Model for Panel Data*

Daniel J. Henderson†
Department of Economics
State University of New York at Binghamton

Léopold Simar‡
CORE and the Institute of Statistics
Université catholique de Louvain, Louvain-la-Neuve

September 1, 2005

Abstract

In this paper we estimate the frontier and time variant technical efficiency fully nonparametrically by exploiting recent advances in kernel regression estimation of categorical data. Specifically, we model firm (unordered) and time (ordered) categorical variables directly into the conditional mean. This approach allows us to smooth the firm and time specific effects, which formally entered the model linearly. Our setup allows for more flexible and accurate estimates of the frontier and time variant technical efficiency. Further, the estimators are consistent and achieve the standard nonparametric rate of convergence. We apply these techniques to a data set examining labor efficiencies of 17 railway companies over a period of 14 years. Not only are our results for the elasticites more economically intuitive than the parametric and semiparametric procedures, we obtain different rankings in terms of labor efficiencies.

Keywords: Technical Efficiency, Nonparametric Kernel, Panel Data
JEL Classification: C1, C14, C33

---

*The authors would like to thank Christopher Parmeter for useful comments on the subject matter of this paper as well as participants of the 9th European Workshop on Efficiency and Productivity Analysis, Brussels, Belgium (June 2005). The research support from "Interuniversity Attraction Pole," Phase V (No. P5/24) from the Belgian Government (Belgian Science Policy) is also gratefully acknowledged.

†Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY 13902-6000, U.S.A., (607) 777-4480, Fax: (607) 777-2681, e-mail: djhender@binghamton.edu.

‡Léopold Simar, Center for Operations Research and Econometrics and the Institute of Statistics, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium, +32 10 47.43.08, Fax: +32 10 47.30.32, e-mail: simar@stat.ucl.ac.be.

1 Introduction

In recent years, econometric research has set to relax many of the restrictive assumptions used to estimate economic models. The area of efficiency measurement is no exception. Beginning with Schmidt and Sickels (1984), who relaxed distributional assumptions on the error components, econometricians have attempted to further relax the parametric assumptions in production frontier models. Here the ongoing debate has been between using nonparametric methods which use linear programming methods versus parametric methods which use econometrics. Although the debate continues, it is obvious that a stochastic procedure which adopts nonparametric techniques would be optimal. The use of nonparametric kernel methods has helped facilitate this trend in the literature (e.g., see Adams, Berger and Sickles 1999, Fan, Li and Weersink 1996, Kneip and Simar 1996, Park, Sickles and Simar 1998, 2003 and Sickles 2005).

Evidence for this suggestion comes from the Monte Carlo exercises of Gong and Sickles (1992). They show that estimates of technical efficiency for parametric panel data stochastic frontier models are improved when the employed form of the production function is closer to the true underlying technology. In general, when a data generating process is unknown, nonparametric kernel methods often give the most reliable results (of course, the relative performance also depends on such things as dimensionality and sample size).

The objective of this paper is to provide fully nonparametric estimators of production frontiers and time variant technical efficiency. Based on recent results on regression with continuous and categorical data (Racine and Li 2004) we improve upon past attempts by allowing firm and time effects to be smoothed as well as the continuous regressors. This approach makes the estimation procedure fully nonparametric and thus only requires minimal restrictions on the technology. Further, we see that the rate of convergence of the estimators are of the standard nonparametric rate. We feel that these procedures should prove fruitful for estimating observation specific estimates of both elasticities and technical efficiencies in a panel data setting.

The remainder of this paper is organized as follows: Section 2 provides a brief explanation of past measures, whereas the third section defines our model, proposes the nonparametric estimators and gives some asymptotic results. Section 4 provides an empirical example. Finally, Section 5 concludes the paper.

2 Past Modeling Strategies

In this section we describe previous methods used for estimating production frontiers and technical efficiency in a panel data setting. In these models the production frontiers are estimated, and time invariant estimates of output oriented technical efficiency for each firm are obtained as a by-product of the exercise. This basic framework assumes that we observe a cross-section of data on N firms over T time periods. Quantities of d inputs are used to produce a scalar output through a production function. More specifically, the production frontier model can be written as

$$y_{it} = f(x_{it}, \beta) \exp(\varepsilon_{it} - u_i), \tag{1}$$

where y_{it} represents the level of output for firm i at time period t, f is the production frontier, β is a $d \times 1$ vector of unknown parameters, ε_{it} represents the two sided noise component, and u_i is the non-negative technical inefficiency component of the error term. Estimation of the production frontier, as well as the technical efficiency term can be estimated in several ways. For the remainder of this section we will focus on two specific estimation techniques: one parametric, where the functional form of f will be assumed, and a second semiparametric, where we will allow the data to tell us the form of the production frontier.

2.1 Parametric Estimation

Although many methods exist for estimating a parametric model with panel data, here we choose one of the most popular methods, which is comparable to the remaining procedures. Fixed Effects (FE) estimation of the production frontier, introduced by Schmidt and Sickles (1984), can be obtained, for example, from the log-linear Cobb-Douglas one-way error component model

$$y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}, \tag{2}$$

where y_{it} again represents the output for firm i at time period t. x_{it} is the $d \times 1$ vector of inputs, β is the d dimensional vector of parameters, α_i $(= \alpha - u_i)$ is the firm fixed effect and ε_{it} is the random disturbance. In other words, we assume that each firm shares the same parametric technology in each time period, but that differences between them are captured by a location (firm) effect α_i (time effects can also be accounted for, for example, by simply detrending the data). Estimation of β (elasticities) can be obtained, for example, by means of the within estimator. Firm specific estimates of α are then obtained by

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - x_{it}\widehat{\beta} \right). \tag{3}$$

Here we note that the estimate of β is consistent as $NT \to \infty$, but $\widehat{\alpha}_i$ is consistent only if $T \to \infty$. Further, we estimate the individual u_i by means of the normalization

$$\widehat{u}_i = \max_i \widehat{\alpha}_i - \widehat{\alpha}_i,$$

and firm specific estimates of technical efficiency are given by

$$\widehat{TE}_i = \exp\left(-\widehat{u}_i\right).$$

2.2 Semiparametric Estimation

The (linear) parametric assumption (above) may not be suitable for all panel data sets. If one assumes a linear specification and the Data Generating Process (DGP) is non-linear, then the estimates will most likely be biased. To counter situations such as this, Kneip and Simar (1996) suggest a more general form for the production function:

$$y_{it} = h(x_{it}) + \alpha_i + \varepsilon_{it}, \tag{4}$$

where $h(\cdot)$ is an unknown smooth production function that each firms shares, but differences between them are captured by the location effect α_i. The DGP leading to this model will be described in details in the next section, as a particular case of a more general nonparametric model. There, $h(x)$ will be viewed as an "average" production function, where the average is over the population of firms. Consequently, the "average" of α_i is supposed to be equal to 0 (see Section 3.2 for details).

Estimation of h can be obtained several ways. Kneip and Simar (1996) propose using a Nadaraya-Watson type estimator, but here we suggest using the Local Linear Least Squares (LLLS) estimator. Not only does the local linear estimator give more efficient estimates of $h(\cdot)$, it also allows for estimation of both the production function and the elasticities in one step. To use this method, one must take the first order Taylor expansion of $h(x_{it})$ at the point x, where x is any value of interest in the range of the inputs. This leads to

$$y_{it} \approx h(x) + (x_{it} - x)\beta(x) + \alpha_i + \varepsilon_{it}, \tag{5}$$

where $\beta(x) \equiv \bigtriangledown h(x)$. Thus, when x and y are expressed in logarithmic form, $\beta(x)$ is interpreted as an elasticity. Estimation of $\delta(x) \equiv (h(x), \beta'(x))'$ is then obtained as

$$\widehat{\delta}(x) \equiv \left(\widehat{h}(x), \widehat{\beta}'(x)\right)' = \left(X'K(x)X\right)^{-1} X'K(x)y, \tag{6}$$

where $X = (1, (x_{it} - x))$ and $K(x)$ is a $NT \times NT$ diagonal matrix of kernel (weight) functions $K\left(b^{-1}\left(x_{it} - x\right)\right)$ (note that generally kernel functions can be any probability function having a finite second moment) and b are the optimal bandwidths (known to be the most important factor in kernel estimation). Estimation of the bandwidths can be obtained by using the Least-Squares Cross-Validation (LSCV) procedure. In short, the procedure chooses b such that it minimizes the LSCV function given by

$$CV(b) = \sum_{t=1}^{T} \sum_{i=1}^{N} [y_{it} - \widehat{h}_{-i}(x_{it})]^2, \tag{7}$$

where $\widehat{h}_{-i}(x_{it})$ is computed by leaving out, as in Kneip and Simar (1996), all the T observations (y_{it}, x_{it}) of the ith firm (leave-one firm-out estimator).

Having determined an estimator $\widehat{h}$ of h, estimators of α (again, time effects can also be accounted for, for example, by simply detrending the data) for each i are also obtained by the method of least squares as

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \widehat{h}(x_{it}) \right). \tag{8}$$

Under the regularity assumptions of Kneip and Simar (1996), as $NT \to \infty$

$$|\widehat{h}_i(x) - h_i(x)| = O_p\left((NT)^{-2/(d+4)} + N^{-(1/2)} \right),$$

where $h_i(x) = h(x) + \alpha_i$. Further, under the same regularity conditions, when $N \to \infty$

$$|\widehat{\alpha}_i - \alpha_i| = O_p\left((NT)^{-2/(d+4)} + T^{-(1/2)} + N^{-(1/2)} \right).$$

Again, we estimate the individual u_i by means of the normalization

$$\widehat{u}_i = \max_i \widehat{\alpha}_i - \widehat{\alpha}_i,$$

and firm specific estimates of technical efficiency are given by

$$\widehat{TE}_i = \exp\left(-\widehat{u}_i\right).$$

3 A More General Approach

Although the semiparametric approach above was a step in the right direction, it has several restrictive assumptions. First, the firm effect enters in linearly. Second, it is assumed to be time invariant. Finally, although one can detrend the data on inputs and output, these types of procedures are parametric in nature and do not allow for optimal smoothing.

In this section we propose our fully nonparametric model. Then we will show how to estimate the model as well as give the rates of convergence for our estimators. Our results will show that our estimation procedure is appealing and, as intuitively expected, much more flexible than the preceding ones.

3.1 A Nonparametric Model

A general nonparametric model can be written as follows:

$$y_{it} = h(x_{it}, i, t) + \varepsilon_{it}, \tag{9}$$

where y_{it} is output and $x_{it} \in \mathbb{R}^d$ is a vector of inputs for firm i at time period t, $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Here we will address the problem of estimating the production function, $h(x, i, t)$ which is allowed to vary over each firm and each time period, for all x in a compact subinterval $D \subset \mathbb{R}^d$.

Here we first describe the DGP by introducing some regularity conditions on the stochastic elements of the model, specifically adapting from those given in Kneip and Simar (1996). The setup is quite general and very flexible. First, we consider t as non-random and the value of T will be held fixed (as is the situation in many panel models, the asymptotic case for $T \to \infty$ does not have any practical meaning). Then we specify the stochastics of the DGP through the following assumptions:

- (A1) *Production function and random sample of firms*: for a fixed firm i and a given time period t, the production function is given by $h(x, i, t)$, each firm i is independently drawn from a population (real or conceptual) of firms. Therefore, for each t, $h(\cdot)$ are i.i.d. random functions. Furthermore, we assume each possible realization of the random function $h(\cdot)$ is a smooth function in x (at least twice continuously differentiable).

- (A2) *Frontier function*: for any $x \in D$, the functional values $h(x, i, t)$ are i.i.d. real valued random variables with unknown density $\phi_{x,t}$. The frontier function at time period t (the geometrical locus of optimal production plans) is defined as follows:

  $$f_t(x) = \sup\{y \in \mathbb{R} | \phi_{x,t}(y) > 0\},$$

  and we assume that $f_t(x) < \infty$ for all $x \in D$.

- (A3) *Error term*: the error term ε represents the stochastic component in (9). It is the usual random noise term. We make the usual standard assumption that the ε_{it} are i.i.d. zero mean random variables.

- (A4) *Inputs*: for a given t, the x_{it} are independent continuous d-variate inputs with an unknown density ρ_{it}, with support D. Thus, we allow for some time dependence between the inputs. Note that the results of Li and Racine (2005) show that we can allow for weak dependence between the x_{it}.

– (A5) *Independence of stochastic components*: h, x_{it} and ε_{it} are independent.

These regularity conditions are similar to those used in Kneip and Simar (1996) who also propose a nonparametric estimator of a special case of our model, where $h(x,i,t) = h_i(x)$ for all $t = 1, \ldots, T$. Since the production function is time invariant, they propose a nonparametric estimator of $h_i(x)$ based on a standard kernel estimator (Nadaraya-Watson) obtained by averaging the weighted (by the kernel) values of y_{it} over the T periods of time. The drawback of this approach (as pointed by Kneip and Simar) is that they need large values of T to get sensible estimators of the production functions and they do not take advantage of the information on the production process across the different firms.

For this reason, Kneip and Simar suggest the use of their semiparametric model given in (4) where they restrict the nonparametric $h_i(x)$ to the previously defined additive structure ($h_i(x) = h(x) + \alpha_i$). Thus, the nonparametric estimation of $h(x)$ can be taken by smoothing over all NT observations. In the general approach we propose here, we will consider the nonparametric model where the production frontier can vary over time. We will take advantage of the information over different time periods and across different firms by smoothing the values of y_{it} not only over the values of x_{it} but also over the firm and the time index (i,t); so we are also using the NT observations to estimate $h(\cdot)$. To achieve this, we adapt to this panel situation the method proposed in Li and Racine (2004) and Racine and Li (2004) for regression with both categorical and continuous regressors.

The semiparametric restriction of our model, described in section 2.2 and analog to (4) would be:

$$h_{SP1}(x,i,t) = \alpha_i + h_t(x), \tag{10}$$

where $h_t(x) = E(h(x,i,t)))$, the expectation being relative to $\phi_{x,t}$, so that $E(\alpha_i) = 0$.

Note that an alternative semiparametric model would be a fully additive effect model where we consider a further decomposition of $h_t(x)$:

$$h_{SP2}(x,i,t) = \alpha_i + \gamma_t + h(x), \tag{11}$$

where $h(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$, so that $\sum_{t=1}^{T} \gamma_t = 0$. This latter model will not be developed further in this paper.

We discuss below how to estimate the production function $h(x,i,t)$ for any $x \in D$ and we also particularize to the semiparametric model (10) and provide a nonparametric estimator of $h_t(x)$ in this case.

3.2 A Nonparametric Estimator

Recent advances in nonparametric econometrics allow us to also smooth firm effects (as well as other categorical variables – both unordered and ordered). Here we use Li-Racine Generalized Kernel Estimation (Li and Racine 2004; Racine and Li 2004) to estimate our model (9). Taking a first-order Taylor expansion of $h(x_{it}, i, t)$ with respect to x_{it} at a point x, where x is any value of interest in the range of the inputs, yields

$$y_{it} \approx h(x, i, t) + (x_{it} - x)\beta(x, i, t) + \varepsilon_{it} \tag{12}$$

where $\beta(x, i, t)$ is defined as the partial derivative of $h(x, i, t)$ with respect to x. Again, if y and x are both expressed in logarithmic form, then $\beta(x, i, t)$ is interpreted as a "local" elasticity for the firm i, at time t, evaluated at the point x.

The estimator of $\delta(x, i, t) \equiv (h(x, i, t), \beta'(x, i, t))'$ is given by

$$\begin{aligned}
\widehat{\delta}(x, i, t) &= \left(\widehat{h}(x, i, t), \widehat{\beta}'(x, i, t)\right)' \\
&= \left[\sum_{j,s} K_{(b,\lambda^u,\lambda^o)} \begin{pmatrix} 1 & x_{js} - x \\ x_{js} - x & (x_{js} - x)(x_{js} - x)' \end{pmatrix}\right]^{-1} \\
&\quad \times \sum_{j,s} K_{(b,\lambda^u,\lambda^o)} \begin{pmatrix} 1 \\ x_{js} - x \end{pmatrix} y_{js},
\end{aligned} \tag{13}$$

where $\sum_{j,s}$ stands for $\sum_{j=1}^{N}\sum_{s=1}^{T}$ and $K_{(b,\lambda^u,\lambda^o)} = \ell^c \times \ell^u \times \ell^o$ is the commonly used product kernel function (e.g., see Pagan and Ullah 1999). Here ℓ^c is the standard kernel function used for each of the continuous inputs as described in the previous section with window width b. Moreover ℓ^u is a variation of Aitchison and Aitken's (1976) kernel function for unordered categorical variables, which equals one if $j = i$ and λ^u otherwise and ℓ^o is the Wang and Van Ryzin (1981) kernel function for ordered categorical variables which equals one if $s = t$ and $(\lambda^o)^{|t-s|}$ otherwise. So the weighting in the localizing process does not only depend (as usually when smoothing over continuous covariates) on the distance between x and the data points x_{js} but also on the difference between the firm i and the time t indices with the data j, s indices.

As noted previously, estimation of the bandwidths $(b, \lambda^u, \lambda^o)$ is typically the most salient factor when performing nonparametric estimation. The extension of the LSCV procedure here is trivial. Now the procedure chooses $(b, \lambda^u, \lambda^o)$ which minimize the LSCV function given by

$$CV(b, \lambda^u, \lambda^o) = \sum_{t=1}^{T}\sum_{i=1}^{N}[y_{it} - \widehat{h}_{-i}(x_{it}, i, t)]^2, \tag{14}$$

where $\widehat{h}_{-i}(x_{it}, i, t)$ is computed, as in Kneip and Simar (1996), by leaving out all the T observations (y_{it}, x_{it}) of the ith firm (leave-one firm-out estimator). In the empirical section below we explain how this can be done in practice.

According to the properties established for nonparametric estimation of regression with both categorical and continuous regressors, our nonparametric estimator $\hat{h}(x, i, t)$ achieves the standard nonparametric rate of convergence $\left(N^{2/(d+4)}\right)$ when $N \to \infty$ (remember that T is fixed) under the appropriate regularity conditions and technical assumptions on the bandwidths, as described in Theorem 2.1 of Li and Racine (2004). Under these conditions, we can write, for a firm i, for any $x \in D$ and any time period t, as $N \to \infty$

$$|\hat{h}(x, i, t) - h(x, i, t)| = O_p((NT)^{-2/(d+4)}), \tag{15}$$

where of course the last term is in fact $O_p(N^{-2/(d+4)})$, since T is fixed, but the writing in (15) exposes the fact that the NT observations are used in the estimation.

3.3 Frontier and Efficiency Estimation

From the definition of the production function above, we propose, as in Kneip and Simar (1996), the following estimator of f_t:

$$\hat{f}_t(x) = \max_{i=1,\ldots,N} \hat{h}(x, i, t), \; t = 1, \ldots, T. \tag{16}$$

Following the arguments in Theorem 2 of Kneip and Simar (1996), it can be seen that the rate of convergence of the frontier level relies on the rates obtained for the production function, with a factor involved by the max operator. It is easy to see that the production frontier in (16) is defined as the maximum production possible across each of the N firms at time period t for each x. We then have, under the same regularity conditions, when $N \to \infty$:

$$|\hat{f}_t(x) - f_t(x)| = O_p((NT)^{-2/(d+4)} \log N + N^{-1}).$$

Once the frontier is defined, we can estimate the efficiency of each firm in each time period. Previous attempts to measure efficiency have consisted of finding an average over time of the difference between the observed output and estimated output (in other words, residuals). These average differences are then used to obtain time invariant technical efficiency through a normalization procedure. However, this type of procedure requires that one firm be deemed technically efficient and that all others be measured against that firm regardless of their level of production. Here, not only will we be able to estimate time variant technical efficiency, but we will actually allow for the possibility that more than one firm is deemed to be efficient in any given time period. Further, the estimation procedure also allows for the possibility that no firm is deemed efficient in any time period.

Specifically, the estimation of time variant technical efficiency goes as follows: in a given time period t, a firm i uses a given amount of inputs x_{it} to produce a given amount of output y_{it} observed with noise ε_{it}. We have shown above how to distinguish the actual (estimated) output $\hat{h}(x_{it}, i, t)$ from the observed output y_{it} and the random noise ε_{it}. The question of how to determine the efficiency for a particular firm at a particular time period is thus the following: given the amount of inputs used by firm i in time period t, could any of the other firms have produced more output than did firm i in that same time period t, using the same amount of inputs x_{it}?

From (13) we can see that the estimates of h can be evaluated at any $x \in D \subset \mathbb{R}^d$. Therefore, we simply evaluate the function for each firm j in the sample in that same time period t, using the input levels x_{it} that were used by firm i. Essentially the measure of technical efficiency compares the difference between the actual (estimated) output of firm i and the maximal potential output produced by any other firm in the sample for the same time period. If the output is measured in logarithms, then the estimate of technical efficiency for firm i in time period t is defined as

$$TE_{it} = \frac{\exp\left(\hat{h}\left(x_{it}, i, t\right)\right)}{\exp\left(\max_{j=1,\ldots,N}\left(\hat{h}\left(x_{it}, j, t\right)\right)\right)}.$$

This measure of technical efficiency is simply the ratio of actual to potential output.

In order to help visualize this measure, consider the following trivial example shown in Figure 1. Here we have a sample of two firms (1,2) in a particular time period (t) with a single input (x) and a single output (y). Each firm in period t sits on its production function, $h(x, 1, t)$, and $h(x, 2, t)$ respectively. The production frontier at time period t ($f_t(x)$) is defined as the locus of optimal production plans. It is easy to see, that up to an input level around 35, firm 1 defines the production frontier, however after that input level, firm 2 defines the production frontier. Estimation of efficiency is also straight forward from the figure. Suppose the observed input for firm 1 at time period t was 80. The actual (estimated) output for firm 1 ($h(80, 1, t)$) is roughly equal to 22. However, employing that same amount of input, firm 2 could produce ($h(80, 2, t)$) roughly 30 units of output. Therefore the efficiency for firm 1 at time period t would be the ratio of actual (estimated) to potential output, 0.733 ($= 22/30$).

4 Empirical Example

In this section we will use an empirical example to illustrate the above procedures. Here we consider the analysis of labor efficiencies of 17 railway companies over a period of 14 years (annual observations). Although the sample size (238 observations) is relatively small, this example will allow us to show, from a practical point of view, how the procedures work. Data on the activity of the main international railway companies can be found in the annual reports of the Union Internationale des Chemins de Fer (U.I.C.). The railways retained over the period 1970-1983 can be found in Table 1. One reason for choosing this data set is that it has also been studied in a similar fashion by Hall, Härdle and Simar (1995), Kneip and Simar (1996), and Park and Simar (1994).

Since we analyze a labor function, the support of h is bounded from below (the most technically efficient railway company uses the least labor), and thus the firm with the smallest value of α will be deemed the most efficient in the sample. Following Kneip and Simar (1996), the variables used in this example are

y_{it} = labor (total number of employees) / total length of network (in kms)

x_{1it} = total distance covered by trains (in 10^3 kms) / total length of network (in kms)

x_{2it} = ratio of passenger trains in x_{1it} (in %)

x_{3it} = density of network (kms of lines by 100 km^2).

All the variables are in logarithms and for the first two models (parametric and semiparametric) have been adjusted for the time trend effect. Note that Kneip and Simar (1996) have placed in the denominator of both y and x_1 the total length of the network which eliminates the size effect. Further, the variable x_1 represents a rough measure of the output (demand) of the railways, whereas x_2 characterizes some aspects of the demand and x_3 is a physical measure of the density of the network.

Here we present the results for the three separate models. First, we provide the results for the estimation of the log-linear model

$$y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it},$$

where $x_{it} = (x_{1it}, x_{2it}, x_{3it})$. The top part of Table 2 gives the estimates for the elasticities from the within estimator as well as the associated standard errors (in italics). Note that the last two slopes of the linear model have negative signs and the latter is significant (these are unexpected signs – a higher percentage of passenger trains and an increase in the density of the network should lead to an increased demand for labor). At the same time, the first column of Table 3 gives the estimates of time invariant technical efficiencies of each network. The estimated values of technical efficiency and the ranking of the networks (seen in Table 4) rely on the restrictive linear hypothesis. The next logical question is then, what happens to the estimates when a less restrictive model on the technology is estimated?

Kneip and Simar (1996) suggest the semiparametric model

$$y_{it} = h(x_{it}) + \alpha_i + \varepsilon_{it}$$

to estimate the common labor function and associated firm effects. Here, h is the unknown smooth labor function common to all firms and α_i again represent the individual effect of the ith firm. Further, since we are also interested in estimating the returns to each input ($\beta(x) \equiv \bigtriangledown h(x)$) in a single step, we employ the LLLS estimator. This gives us observation specific estimates of the labor function. The middle part of Table 2 gives the mean values of the LLLS estimates of the parameters. The optimal bandwith was selected by cross-validation as explained above. For the continuous multivariate variable x, we define the bandwidths as $b_j = b_{basis} \times \sigma_{x_j} \times (NT)^{-1/(4+d)}$, for $j = 1, 2, 3$ where the optimal value for b_{basis} is determined by the cross-validation criterion. The optimal value was found to be $b_{basis} = 0.83$. The table also gives the bootstrapped (500 bootstrap replications) standard errors of the estimates. Note now that the first and third coefficients are positive, but insignificant at the means (again, the sign of the coefficient on x_2 at the mean, although insignificant, is unexpected).

In order to compare across two or more models, a unit-free goodness-of-fit measure is usually used. Given the known drawbacks of the R^2 measure based on the decomposition of the sum of squares, we employ an alternative definition that is commonly used. This definition defines R^2 as the squared correlation coefficient between the dependent variable (y) and fitted value ($\widehat{y}$), and is given by

$$R^2 = \rho^2_{y\widehat{y}} = \left[\frac{cov(y, \widehat{y})}{\sqrt{var(y) \times var(\widehat{y})}} \right]^2,$$

which is identical to the standard measure when the model is linear and includes an intercept term. The value given by the parametric model is 0.675 whereas the value for the semiparametric model is 0.892.

The second column of Table 3 also gives the estimates of technical efficiency for each network. Notice the discrepancy between the ranking of the semiparametric case and the ranking of the linear case for a majority of the networks. This information, along with the goodness of fit measure, may lead one to suggest that the parametric assumption is incorrect.

Although this model is less restrictive than the parametric model, the estimated technical efficiencies are assumed to be time invariant. Further, the model is semiparametric ($h(x) + \alpha_i$) and the ranking of the networks rely on these restrictive assumptions. The question now becomes, what happens to the estimates when an even less restrictive model for the technology is estimated?

Our model is thus defined as

$$y_{it} = h(x_{it}, i, t) + \varepsilon_{it},$$

where h is an unknown smooth labor function. Note that the independent and dependent variables are no longer adjusted for the time trend effect because this will be picked up by the categorical variable t. We estimate the model by the procedure explained in Section 3.2. Here three optimal bandwidths have to be selected simultaneously. For the continuous variables we proceed as in the preceding case and for the unordered and ordered categorical variables we follow the Li and Racine (2004) and Racine and Li (2004) scaling suggestions (scaling the bandwidths by a factor $(NT)^{-2/(4+d)}$). Finally, the procedure provides the following optimal values: for the continuous variables, $b_j = 34.00 \times \sigma_{x_j} \times (NT)^{-1/(4+d)}$, $j = 1, 2, 3$ and for the categorical bandwidths, $\lambda^u = 1.81 \times (NT)^{-2/(4+d)}$ and $\lambda^o = 0.81 \times (NT)^{-2/(4+d)}$ The bottom line of Table 2 provides mean values of the parameter estimates as well as their associated bootstrapped (500 bootstrap replications) standard errors. Note that with the nonparametric approach the returns take their expected signs and that the first and third are significant. Further, we see an improvement in the goodness-of-fit measure with a $R^2 = 0.954$.

The last column of Table 3 gives the average (for the nonparametric case) technical efficiency for each network. Note that although some of the rankings are similar with the nonparametric and semiparametric procedures (e.g., ÖBB), many of the rankings differ significantly (e.g., DSB).

Of course in this approach we have more information on the firm and time dependent structure of the efficiencies. Table 5 provides all the efficiency scores for each firm at each time period with the corresponding averages. For instance, we see that five firms are deemed efficient in at least one time period (CP, DB, NS, SNCB and SJ). Further, we see that NS and SJ are considered to be efficient in each time period. Thus, neither of these two firms dominates the other. Of course, this type of detailed analysis is not availbale with the other techniques described above.

Also notable from this table, is the fact that the time invariant assumption does not hold for the technical efficiencies. These restrictions (especially the time invariant assumption), in the semiparametric framework, may penalize some networks. For example, if we simply look at the average technical efficiency for each firm, the results for CH look similar between the nonparametric and semiparametric measures. However, by examining Figure 2 we see that although the two measures were similar at the beginning of the sample, towards the end of the sample, the technical efficiency falls in the nonparametric case. As another example, consider the case of JNR versus VR. In terms of average efficiency, the two networks are ranked last and second to last respectively. Further, the average estimate of technical efficiency for JNR appears to be much less than that of VR. Thus, given these average results, one might believe that JNR is much worse off than VR. However, if we look at Figure 3, which plots technical efficiencies over time we can see that although JNR had a relatively poor efficiency score at the beginning of the sample, as we approach the end of the sample, JNR improves in terms of efficiency. At the same time, VR had a significantly higher efficiency score at the beginning of the sample, but in fact JNR slightly overtakes VR in terms of efficiency scores during the final period.

In our illustration with railway companies we have shown the problems induced by using restrictive assumptions on the technology when it could be incorrect. This might provide spurious inefficiencies and wrong efficiency scores which penalizes some networks.

5 Conclusion

This paper presents a procedure to estimate fully nonparametric production frontiers as well as estimate time variant measures of technical efficiency. This is achieved by exploiting recent advances in econometrics which allowed us to smooth categorical variables. Specifically, we used Generalized Kernel Estimation to smooth both the continuous regressors as well as the categorical regressors representing both individual firms and time periods. This added flexibility did not come at an additional cost because we found that the estimates of both the production function and the production frontier converge at the standard nonparametric rate.

The illustration to a real data set indicates that the procedure is easy to implement and provides sensible results, much more sensible than those obtained by more restrictive models. Our results have shown that the estimation procedure is appealing and, as intuitively expected, much more flexible than the other ones. The analysis provides also more insights into the structure and the evolution of efficiencies over time and across firms.

References

[1] Adams, R.M., A.N. Berger and R.C. Sickles (1999). "Semiparametric Approaches to Stochastic Panel Frontiers with Applications in the Banking Industry," *Journal of Business and Economic Statistics*, 17, 349-58.

[2] Aitchison, J. and C.B.B. Aitken (1976). "Multivariate Binary Discrimination by Kernel Method," *Biometrika*, 63, 413-420.

[3] Fan, Y. Q. Li and A. Weersink (1996). "Semi-parametric Estimation of Stochastic Production Frontier Models," *Journal of Business and Economic Statisitics*, 14, 460-8.

[4] Gong, B.-H. and R.C. Sickels (1992). "A Comparison Between Stochastic Frontier and Data Envelopment Methods Using Panel Data," *Journal of Econometrics,* 51, 259-84.

[5] Hall, P., W. Härdle and L. Simar (1995). "Iterated Bootstrap with Application to Frontier Models," *Journal of Productivity Analysis,* 6, 63-76.

[6] Kneip, A. and L. Simar (1996). "A General Framework for Frontier Estimation with Panel Data," *Journal of Productivity Analysis,* 7, 187-212.

[7] Li, Q. and J. Racine (2004). "Cross-Validated Local Linear Nonparametric Regression," *Statistica Sinica*, 14, 485-512.

[8] Li, Q. and J. Racine (2005). *Nonparametric Econometrics: Theory and Practice,* Princeton, Princeton University Press, forthcoming.

[9] Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge, Cambridge University Press.

[10] Park, B., R.C. Sickels and L. Simar (1998). "Stochastic Frontiers: A Semiparametric Approach," *Journal of Econometrics*, 84, 273-301.

[11] Park, B., R.C. Sickels and L. Simar (2003). "Semiparametric Efficient Estimation of AR(1) Panel Data Models," *Journal of Econometrics*, 117, 279-309.

[12] Park, B. and L. Simar (1994). "Efficient Semiparametric Estimation in a Stochastic Frontier Model," *Journal of the American Statistical Association,* 89, 929-36.

[13] Racine, J. and Q. Li (2004), "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics,* 119, 99-130.

[14] Schmidt, P., and R.C. Sickles (1984). "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics* 2, 367-74.

[15] Sickles, R.C. (2005). "Panel Estimators and the Identification of Firm-Specific Efficiency Levels in Parametric, Semiparametric and Nonparametric Settings," *Journal of Econometrics*, 126, 305-34.

[16] Wang, M.C., and J. Van Ryzin (1981), "A Class of Smooth Estimators for Discrete Estimation," *Biometrika*, 68, 301-309.

[17] U.I.C. (1970-1983). *Statistiques Internationales des Chemins de Fer,* Union Internationale des Chemins de Fer, Paris.

Table 1 - Railway Companies

Network	Country
BR	Great Britain
CFF	Switzerland
CFL	Luxemburg
CH	Greece
CP	Portugal
DB	Germany
DSB	Denmark
FS	Italy
JNR	Japan
NS	Netherlands
NSB	Norway
ÖBB	Austria
RENFE	Spain
SJ	Sweden
SNCB	Belgium
SNCF	France
VR	Finland

Table 2 - (Average) Elasticity Estimates

	x_1	x_2	x_3
Parametric	0.111	-0.096	-0.637
	0.068	*0.138*	*0.141*
Semiparametric	0.079	-0.261	0.082
	0.222	*0.258*	*0.072*
Nonparametric	0.333	0.433	0.432
	0.181	*0.396*	*0.110*

Table 3 - (Average) Estimates of Technical Efficiency

Network	Parametric	Semiparametric	Nonparametric
BR	0.122	0.922	0.871
CFF	0.118	0.909	0.805
CFL	0.078	0.913	0.820
CH	0.600	0.900	0.876
CP	0.281	0.902	0.888
DB	0.088	0.914	0.967
DSB	0.206	1.000	0.848
FS	0.129	0.834	0.785
JNR	0.095	0.845	0.614
NS	0.145	0.904	1.000
NSB	1.000	0.902	0.818
ÖBB	0.120	0.853	0.801
RENFE	0.477	0.870	0.887
SJ	0.844	0.944	1.000
SNCB	0.071	0.903	0.992
SNCF	0.208	0.933	0.892
VR	0.811	0.895	0.758

Table 4 - (Average) Efficiency Estimate Rankings

Network	Parametric	Semiparametric	Nonparametric
BR	11	4	9
CFF	13	7	13
CFL	16	6	11
CH	4	12	8
CP	6	10	6
DB	15	5	4
DSB	8	1	10
FS	10	17	15
JNR	14	16	17
NS	9	8	1
NSB	1	10	12
ÖBB	12	15	14
RENFE	5	14	7
SJ	2	2	1
SNCB	17	9	3
SNCF	7	3	5
VR	3	13	16

Table 5 – Observation Specific Efficiency Scores

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	Mean
BR	0.864	0.869	0.887	0.893	0.890	0.879	0.874	0.861	0.853	0.852	0.849	0.856	0.888	0.879	0.871
CFF	0.827	0.830	0.824	0.818	0.813	0.808	0.811	0.814	0.801	0.798	0.803	0.792	0.784	0.754	0.805
CFL	0.801	0.804	0.831	0.836	0.818	0.838	0.835	0.823	0.827	0.806	0.811	0.815	0.822	0.817	0.820
CH	0.906	0.913	0.908	0.907	0.894	0.890	0.879	0.868	0.868	0.857	0.850	0.843	0.836	0.844	0.876
CP	0.792	0.804	0.850	0.861	0.861	0.847	0.835	0.858	0.869	0.907	0.960	1.000	1.000	0.992	0.888
DB	0.935	0.935	0.944	0.948	0.949	0.946	0.947	0.956	0.974	0.999	1.000	1.000	1.000	0.999	0.967
DSB	0.862	0.860	0.827	0.829	0.833	0.836	0.836	0.839	0.842	0.846	0.847	0.839	0.848	0.931	0.848
FS	0.792	0.789	0.785	0.784	0.783	0.774	0.778	0.777	0.773	0.778	0.787	0.790	0.800	0.798	0.785
JNR	0.583	0.590	0.597	0.602	0.603	0.612	0.605	0.602	0.608	0.608	0.622	0.638	0.654	0.668	0.614
NS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NSB	0.820	0.835	0.838	0.838	0.822	0.830	0.835	0.823	0.806	0.793	0.794	0.804	0.806	0.809	0.818
OBB	0.844	0.839	0.834	0.821	0.797	0.809	0.788	0.785	0.793	0.784	0.780	0.778	0.777	0.781	0.801
RENFE	0.854	0.873	0.866	0.877	0.874	0.888	0.851	0.901	0.897	0.899	0.896	0.904	0.897	0.885	0.877
SJ	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SNCB	0.985	0.988	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991	0.978	0.965	0.988	1.000	0.992
SNCF	0.898	0.898	0.898	0.898	0.898	0.899	0.884	0.875	0.874	0.881	0.891	0.901	0.894	0.898	0.892
VR	0.802	0.816	0.809	0.805	0.787	0.799	0.792	0.788	0.773	0.736	0.700	0.678	0.665	0.666	0.758
Mean	0.857	0.861	0.866	0.866	0.860	0.862	0.858	0.857	0.856	0.855	0.857	0.859	0.862	0.866	0.860

Figure 1

Two Firm, One Input, One Output Model

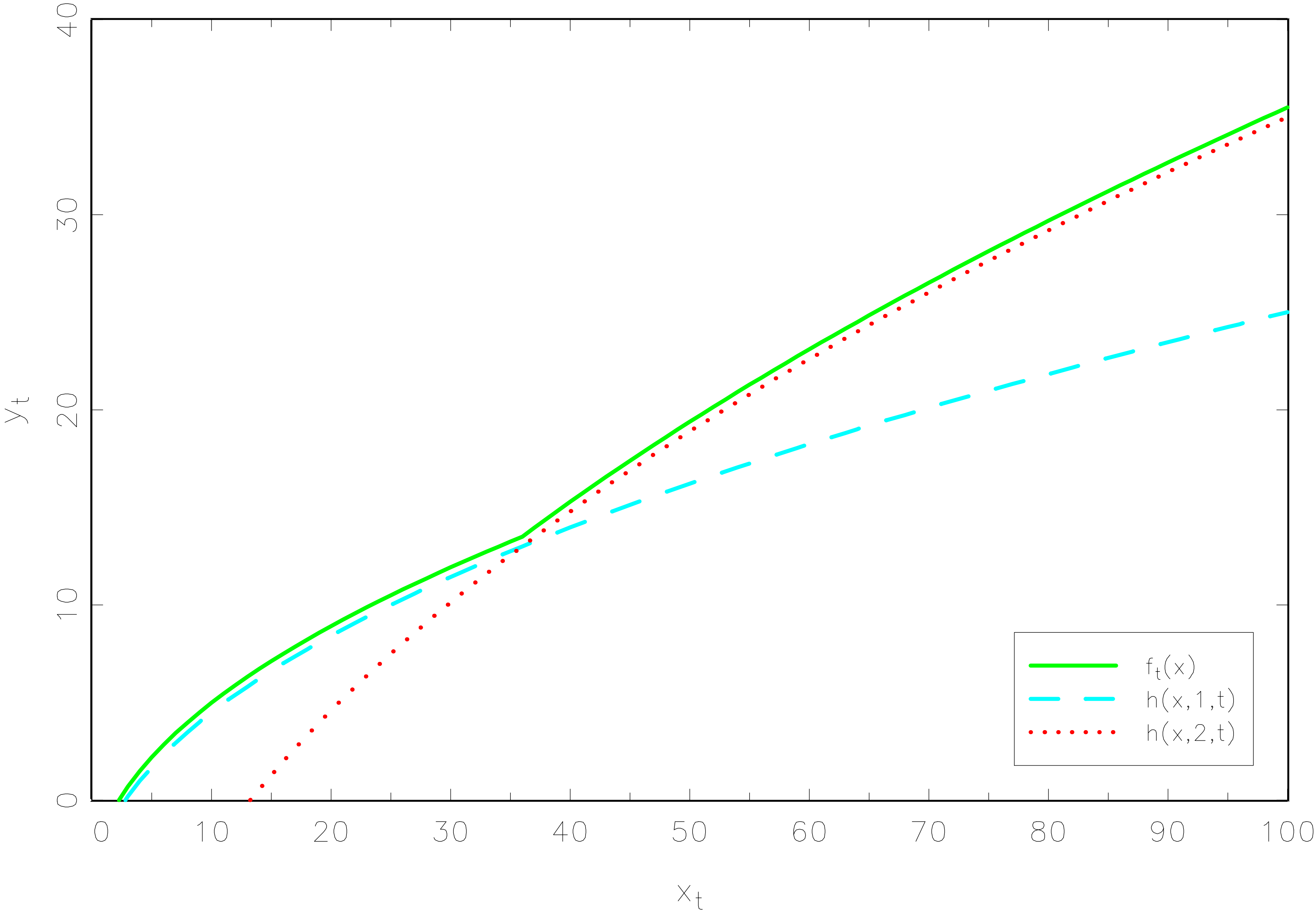

Figure 2
Efficiency Over Time (CH)

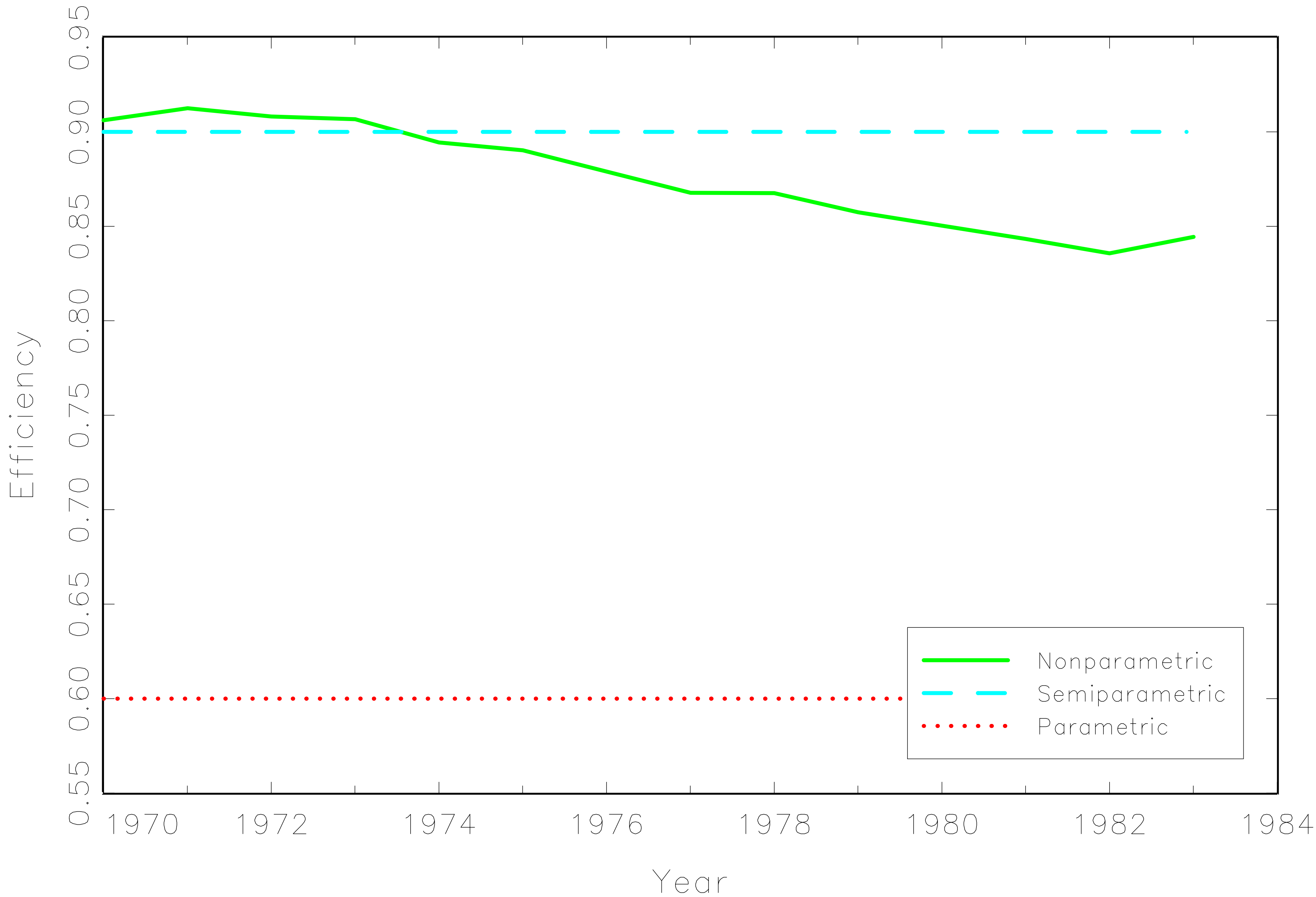

Figure 3
Efficiency Over Time (JNR & VR)