A General Framework for Frontier Estimation with Panel Data

August 1995 [1]

by Alois KNEIP and Léopold SIMAR

CORE and Institut de Statistique
Université Catholique de Louvain
34, voie du roman pays
B-1348, Louvain-la-Neuve, Belgium
Tel: 3210474314, Fax: 3210473032
e-mail: simar@stat.ucl.ac.be; kneip@stat.ucl.ac.be

Abstract

The main objective of the paper is to present a general framework for estimating production frontier models with panel data : a sample of firms $i = 1, \ldots, N$ is observed on several time periods $t = 1, \ldots, T$. In this framework, nonparametric stochastic models for the frontier will be analysed. The usual parametric formulations of the literature are viewed as particular cases and the convergence of the obtained estimators in this general framework are investigated. Special attention is devoted to the role of N and of T on the speeds of convergence of the obtained estimators. First, a very general model is investigated, in this model almost no restriction is imposed on the structure of the model or of the inefficiencies. This model is estimable from a nonparametric point of view but needs large values of T and of N to obtain reliable estimates of the individual production functions and estimates of the frontier function. Then more specific nonparametric firm effect models are presented. In these cases, only NT must be large to estimate the common production function ; but again both large N and T are needed for estimating individual efficiencies and for estimating the frontier. The methods are illustrated through a numerical example with real data.

[1]Invited paper presented at the Advanced Research Workshop on Efficiency Measurements, Odense University, 22-24 may 1995. Research support from the contract "Projet d'Actions de Recherche Concertées" (PARC 93/98-164) of the Belgian Government is gracefully acknowledged. The authors thank Christian Weiner for his help in the computational work.

1 Introduction

In the last two decades, the econometric and statistical literature has extensively investigated the problem of the estimation of efficiencies of production units from frontier models. The basic ideas came from the pioneering work of Farrell (1957). The efficiency of production units is generally measured relative to a production frontier, defined as the geometric locus of optimal production. This frontier may be estimated from a set of observed production units. The original approach of Farrell was based on a nonparametric model (no particular functional form for the frontier function) and Aigner and Chu (1968) proposed a parametric alternative. In those two basic papers, no inference (statistical properties) was available either on the estimated frontier or for the derived measures of efficiencies.

However, from a scientific point of view, the question has to be raised about the statistical significance of the obtained estimators (of the frontier and of the efficiencies of the firms).

The major part of the research on that issue was mainly concentrated on parametric models in two different frameworks. In the deterministic framework, all the observations are supposed to fall on the same side of the frontier and the (one-sided) error term captures only inefficiency (see e.g. Greene (1980)). In the (often more realistic) stochastic framework, a composed error term allows for stochastic noise and for inefficiency (see e.g. Aigner, Lovell and Schmidt (1977)).

For the nonparametric approaches, the estimation techniques have also been developed. The proposed methodologies, like DEA (Data Envelopment Analysis for convex production set, see e.g.: Charnes, Cooper and Rhodes (1978) or Färe, Grosskopf and Lovell (1985) or FDH (Free Disposal Hull allowing non convexities, see Deprins, Simar and Tulkens (1984)) are essentially descriptive tools allowing the analysis of the observed technology. They provide also a ranking of the firms but no inference is available. Recently, some papers were devoted to the statistical aspects of these nonparametric estimators, either by proving their statistical consistency (see Korostelev, Simar and Tsybakov (1995a) for FDH, Banker (1993) for DEA and Korostelev, Simar and Tsybakov (1995b) for DEA and FDH) or analysing their sensitivity by resampling techniques (jacknife or bootstrap, see e.g. Andersen and Petersen (1993), Lovell, Walters and Wood (1995) for applications and Simar and Wilson (1995) for a methodological presentation of the bootstrap principles for nonparametric frontiers).

Till now, it appears that most of the effort in the nonparametric approaches was concentrated on the case of deterministic frontiers where no statistical noise or random disturbance for the data is allowed. The aim of this paper is to propose an approach allowing the statistical analysis of **nonparametric stochastic frontier models** when a panel of data is available.

With panel data, a sample of firms $i = 1, \ldots, N$ is observed on several time periods $t = 1, \ldots, T$. The case of panel data is certainly the most interesting when one is interested to investigate the efficiencies of each firm. Efficiency measures come essentially from the observation of the distance between the estimated frontier and the observed firms ; in most situations, this distance is captured by

a residual. With panel data, several observations of the residuals of each firm are available : this allows for instance to test some structural hypothesis on the efficiencies or to appreciate the statistical significance of the efficiency of each firm.

We know also that in stochastic frontier models, it is difficult to identify (even in parametric models) the efficiency part and the stochastic part of the composed error term. Jondrow, Lovell, Materov and Schmidt (1982) proposed an estimator of the efficiency part based on the conditional distribution of the efficiency conditional on the value of the composed term. In a parametric set-up, the two parts are indeed identifiable (due to all the parametric assumptions on both the frontier function and on the probability distributions) but the statistical properties of the resulting estimates are still uncertain. However, in the general setup of nonparametric models, a panel of data will be necessary to identify the stochastic models.

The parametric approach to frontier models with panel data was proposed by Schmidt and Sickles (1984) for the basic ideas and developed e.g. by Cornwell,Schmidt and Sickles (1990), Battese and Coelli (1992), Kumbhakar (1990) and Lee and Schmidt (1993) where several attempts were made in order to model the inefficiency term (various time dependence strucure,...). Efficient semiparametric estimation of these models is proposed by Park and Simar (1994) and Park, Sickles and Simar (1993) where efficient parametric estimates of a linear model are investigated with an unspecified distribution for the inefficiency term ; further some asymptotics is investigated on the resulting individual efficiency scores and on the level of the frontier itself. A comparison by bootstrap of parametric and nonparametric techniques is provided by Simar (1992).

In this paper, nonparametric stochastic frontier models will be presented in a general framework and some estimators will be proposed. It is a first attempt to analyse this general problem. The statistical properties of the obtained estimators will thus be investigated for simple models under some familiar regularity conditions which facilitate the presentation. No doubt that in further works, some of these conditions (like time independence, constant inefficiency over time, ...) will be relaxed and that more sophisticated models and techniques will be developed.

The paper is organized as follows. Section 2 presents the general framework for estimating stochastic production frontier models with panel data. In this presentation, the parametric formulations are viewed as particular cases. The convergence of the obtained estimators in this general framework are investigated. Special attention is devoted to the role of N and of T on the speeds of convergence of the obtained estimators [2]. Section 3 particularizes the model to a firm effect model which can be viewed as a nonparametric generalisation of the familiar parametric random effect model of the literature. In this particular model, the rate of convergence of the estimators is better than for the preceding case since it typically relies on the total number of observations NT. Section 4 proposes estimators of the measure of efficiency of each firm and investigates their convergence properties even in a general setup. Finally, section 5 illustrates the feasibility of the techniques with a numerical example.

[2]The proofs of the theorems are very technical and will not be reproduced in this paper; they will be found in Kneip and Simar (1995)

2 The general model

2.1 The model

The general model can be written as follows :

$$y_{it} = h_i(x_{it}) + \epsilon_{it} \tag{2.1}$$

where y_{it} is the output and $x_{it} \in \mathcal{R}^d$ is a vector of intputs, $i = 1, \ldots, N$ and $t = 1, \ldots, T$. We will consider the problem of estimating each $h_i(x)$ and the resulting frontier for all x in a compact subinterval $D \subset \mathcal{R}^d$.

Through all the paper, some regularity conditions are needed on the stochastics of the model. They are not very restrictive and fix the general setup of the paper.

General Setup.

(i) In this model, for a fixed firm i, the production function is given by $h_i(x)$, each firm being independently drawn from a population (real or conceptual) of firms. In other words, $\{h_i\}_{i=1,\ldots,N}$ are i.i.d. random functions. Furthermore, each possible realization of the random functions h_i is a smooth function, i.e. at least twice continuously differentiable.

(ii) For any $x \in D$, the functional values $h_i(x)$ are i.i.d. real valued random variables with unknown density ϕ_x. The frontier function $f(x)$ (geometrical locus of optimal production situations) is defined as follows

$$f(x) = \sup\{y \in \mathcal{R} | \phi_x(y) > 0\} \tag{2.2}$$

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

(iii) The error term ϵ represent the stochastic component, it is the usual random noise term. A standard hypothesis on ϵ is : ϵ_{it} are i.i.d. zero mean random variables with a common distribution W satisfying

$$\int exp(\beta y^2) W(dy) < \infty \text{ for some } \beta > 0. \tag{2.3}$$

(Note that the condition on W is, for example satisfied if $W = N(0, \sigma^2)$ or if W possesses compact support.)

(iv) The x_{it} are independent d-variate random variables with unknown density ρ_{it} with support D. There exists $c > 0$ such that $\rho_{it}(x) > c$ for all $x \in D$ and $\rho_{it}(x) = 0$ for all $x \notin D$.

(v) h_i, ϵ_{it} and x_{it} are independent.

Some additional technical conditions will be used in some of the theorems; they are described with full details in Kneip and Simar (1995).

2.2 Estimation of $h_i(x)$

In this section we propose a general nonparametric estimation of the individual production functions $h_i(x)$. For a fixed firm i, it is indeed possible to estimate the function $h_i(x)$ since we have, for this firm, T observations. We can use usual nonparametric estimators to provide the functions $\hat{h_i}(x)$. We propose in this paper a simple, not very sophisticated nonparametric estimator : the Nadaraya-Watson kernel estimator. Another possible estimation method is the use of multivariate local polynomial regression. In fact, multivariate local linear estimators provide more efficient estimators than the Nadaraya-Watson kernel estimator (compare Ruppert and Wand (1994)). However the structure of these estimators is more complicated and rates of convergence are identical.

The kernel estimators are very popular in nonparametric regression (see e.g. Härdle (1990)), they can be motivated as follows. The basic idea behind kernel smoothing is the idea of a local averaging procedure. Since for all x, $h_i(x)$ is the mean of y_i, it is indeed intuitively reasonable to define an estimator of $h_i(x)$ as a local average of the value of y_i. Formally,

$$\hat{h_i}(x) = \sum_{t=1}^{T} w_{it}(x) y_{it} \tag{2.4}$$

where $\{w_{it}(x)\}_{t=1,\ldots,T}$ denotes a sequence of weights which may depend on the whole set $\{x_{it}\}_{t=1,\ldots,T}$. The estimator is often called a *smoother* since the averaging smooths the values of y_{it} through the weight sequence which is tuned by a smoothing parameter. Another way of motivating the local averaging procedure above is that it can be shown as being the local weighted least squares estimate of $h_i(x)$. More precisely, supposing that for all x the weights are positive and sum to one, we have

$$\min_{\theta} \sum_{t=1}^{T} w_{it}(x)(y_{it} - \theta)^2 = \sum_{t=1}^{T} w_{it}(x)(y_{it} - \hat{h_i}(x))^2. \tag{2.5}$$

A simple approach to characterize the weight sequence is to describe the shape of the weights through a density function (*the kernel* K) with a scale parameter that adjusts the size and the form of the weights near x. In the case where $d = 1$, the Nadaraya-Watson kernel estimator is based on the sequence of weights defined by

$$w_{it}(x) = \frac{K(b^{-1}(x - x_{it}))}{\sum_{t=1}^{T} K(b^{-1}(x - x_{it}))} \tag{2.6}$$

The shape of the kernel is determined by K and the size of the weights is tuned by the parameter b, which is called the *bandwidth* ; the choice of the bandwith will balance between the bias and the variance of the estimator and influence the speed of convergence (see below).

In the multidimensional case ($d > 1$), one can use a multidimensional product kernel function

$$K(u_1, u_2, \ldots, u_d) = \prod_{j=1}^{d} K^*(u_j) \tag{2.7}$$

where K^* is a univariate kernel. In order to simplify the notation, the product kernel will still be written as $K(b^{-1} \star (x - x_{it}))$ but here, b is to be understood as a vector of bandwidths b_j, $j = 1, \ldots, d$, b^{-1} stands for the vector of the inverses

$$b^{-1} = (\frac{1}{b_1}, \ldots, \frac{1}{b_d}), \tag{2.8}$$

and $\star$ stands for an elementwise vector product. The product kernel is now defined as follows

$$K(b^{-1} \star (x - x_{it})) = \prod_{j=1}^{d} K^*(\frac{x_j - x_{jit}}{b_j}). \tag{2.9}$$

A different bandwith b_j for each component of x is suitable when these components have a different scale [3].

A common assumption on the stucture of the kernel function $K : \mathcal{R}^d \to \mathcal{R}$ is as follows :

(i) There exists a constant c such that $K(x) < c$ for all $x \in D$.

(ii) $K(x) = 0$ if $x \notin [-1, 1]^d$.

(iii) $\int_{\mathcal{R}^d} K(x)dx = 1$ and $\int_{\mathcal{R}^d} xK(x)dx = 0$.

Note that a reasonable kernel satisfying (i)-(iii) must not necessarily be a product kernel. For $d = 1$ reasonable choices of K satisfying (i)-(iii) are the Epanechnikov kernel

$$K(x) = \frac{3}{4}(1 - x^2)I(|x| \leq 1) \tag{2.10}$$

or the quartic kernel

$$K(x) = \frac{15}{16}(1 - x^2)^2 I(|x| \leq 1) \tag{2.11}$$

Both kernels possess certain optimality properties. Based on a mean squared criterion, it can be shown that the Epanechnikov kernel is optimal in the class of all positive kernels satisfying (iii) (compare Gasser, Müller and Mammitzsch (1985)). The quartic kernel is optimal in the class of all smooth, positive kernel functions (compare Müller (1988)). In the d-dimensional case ($d > 1$), one might rely, as above, on the product kernel

$$K(x) = \prod_{j=1}^{d} K^*(x_j) \qquad x = (x_1, \ldots, x_d) \tag{2.12}$$

[3] Typically, if s_j denote the scale of the j^{th} component of x, the bandwith b_j could be defined as $b_j = s_j * b$, where b is the baseline bandwidth.

where K^* denotes the Epanechnikov or the quartic kernel.

Finally, our kernel estimator of the individual production function $h_i(x)$ is

$$\hat{h_i}^{(1)}(x) = \frac{\sum_{t=1}^{T} K(b^{-1} \star (x - x_{it}))y_{it}}{\sum_{t=1}^{T} K(b^{-1} \star (x - x_{it}))} \tag{2.13}$$

The main advantage of this unsophisticated estimator is its flexibility, it is only required that the h_i is a smooth function in order to get the desired convergence properties. Its main disadvantage is precisely the speed of convergence. In order to be useful the sample size T needs to be large. This is not a severe problem if $d = 1$, since there, reasonable estimators might already be expected for $T = 20$. However, the "minimal" sample size required for reasonable estimators increases **rapidly** with d (see below the section on the rates of convergence). This "curse of dimensionality" is a well known problem in nonparametric regression and density estimation (see e.g. Härdle (1990) and Silverman (1986)). In fact, there is a general agreement that from a practical point of view it does not make sense to use a general nonparametric estimator if d is large[4].

The nonparametric estimators are sensitive to the choice of the bandwidth. In practice, bandwidths might be chosen in a data-adaptative way using cross-validation, i.e. by minimizing the "prediction error"

$$\sum_{i,t} (y_{it} - \hat{h}_{i,-t}^{(1)}(x_{it}))^2 \tag{2.14}$$

where $\hat{h}_{i,-t}$ is computed by leaving out the observation (y_{it}, x_{it}). It is well known that cross-validation provides consistent estimators of the optimal bandwidth in many situations. Alternative automatic bandwidth selectors can be based on generalized cross-validation (Craven and Wabha (1979)) or plug-in methods (see, e.g., Gasser, Kneip and Köhler (1991)).

Note that a consistent estimator of $\sigma^2 = Var(\epsilon_{it})$ is also available. In this general context, it is given by :

$$\hat{\sigma}^2 = \frac{1}{NT} \sum_{i,t} (y_{it} - \hat{h_i}^{(1)}(x_{it}))^2 \tag{2.15}$$

The rate of convergence for $\hat{\sigma}^2$ relies on the rate achieved by $\hat{h_i}^{(1)}$ (see below) and is typically of order $\mathcal{O}_p(N^{-\frac{1}{2}} T^{-\frac{2}{d+4}})$. For the case $d = 1$, estimation of σ^2 might also

[4]Other nonparametric models could be proposed to overcome this difficulty. They will not be detailed in this paper. One might for instance think of Generalized Additive Models (see e.g. Hastie and Tibshirani (1990). There, we need an additional model assumption, it is supposed that the h_i admits the following decomposition

$$h_i(x) = h_{i(1)}(x_1) + h_{i(2)}(x_2) + \ldots + h_{i(d)}(x_d)$$

where the $h_{i(1)}, \ldots, h_{i(d)}$ are univariate smooth functions. This model avoids the "curse of dimensionality : for $d > 1$, estimates possess the same rate of convergence as for $d = 1$. Still T needs to be reasonably large but the main disadvantage is that the estimation method is complicated ; further the above decomposition could be a wrong approximation of the real function.

be based on the methods proposed by Gasser, Stoka and Jennen-Steinwetz (1986) or Hall, Kay and Titterington (1990). In the present context, these estimators of σ^2 for $d = 1$, would achieve the better rate of convergence $\mathcal{O}_p((NT)^{-\frac{1}{2}})$.

2.3 Parametric estimates

Of course, as a particular case of the preceding section, if the following additional parametric assumption is made on the function h_i

$$h_i(x) = X'\beta_i + \beta_{0i} \tag{2.16}$$

for some parameter vector $\beta_i \in \mathcal{R}^d$ and $\beta_{0i} \in \mathcal{R}$, we obtain a standard linear frontier model. If β_i, β_{0i} are estimated by OLS, this yields the estimator of h_i

$$\hat{h_i}^{(2)}(x) = X'\hat{\beta}_i + \hat{\beta}_{0i}. \tag{2.17}$$

The main advantages are here nice rates of convergence, simple estimation and no problem with the dimensionality. On the other side, the disadvantages might be serious : the model might be completely wrong. Note that this linear model is also a particular case of the generalized additive model presented in footnote 4 above.

2.4 Rates of convergence

In this section we provide the rates of convergence of both estimators $\hat{h_i}^{(1)}$ and $\hat{h_i}^{(2)}$. The following Theorem states a well-known result :

Theorem 1 *For every x in the interior of D :*

(a) If $b_j \sim T^{-\frac{1}{d+4}}$, for $j = 1, \ldots, d$, then as $T \to \infty$

$$|\hat{h_i}^{(1)}(x) - h_i(x)| = \mathcal{O}_p(T^{-\frac{2}{d+4}}) \tag{2.18}$$

(b) If the linear model for h_i (2.16) holds, then as $T \to \infty$

$$|\hat{h_i}^{(2)}(x) - h_i(x)| = \mathcal{O}_p(T^{-\frac{1}{2}}) \tag{2.19}$$

This theorem is important since it shows, for the general nonparametric approach, the role of d in the speed of convergence (the "curse of dimensionality"). The rate of convergence of the nonparametric estimator is not far from the familiar rate of convergence of the parametric linear model ($T^{-\frac{1}{2}}$) when d is equal to 1, but decreases rapidly when d increases.

It is well known(see e.g. Stone (1982)) that the estimators to be obtained in the generalized additive models (see footnote 4 above) adopt the rate of convergence of $\mathcal{O}_p(T^{-\frac{2}{5}})$ for any value of d.

In any case, the nonparametric model and its linear particular case need large value of T in order to have sensible estimators of $h_i(x)$. This will also be true for the estimation of the frontier but there also, the number of observed firms N will play a central role as shown in the next section.

2.5 Frontier Estimation

From the definition of the frontier function $f(x)$ above, natural estimates of f come from the estimators of h_i. They are given by

$$\hat{f}^{(1)}(x) = \max_{i=1,\ldots,N} (\hat{h_i}^{(1)}(x)) \tag{2.20}$$

or

$$\hat{f}^{(2)}(x) = \max_{i=1,\ldots,N} (\hat{h_i}^{(2)}(x)) \tag{2.21}$$

The properties of these estimators of the frontier are fixed by the following Theorem.

Theorem 2 *For every x in the interior of D :*

(a) If $b_j \sim T^{-\frac{1}{d+4}}$, for $j = 1, \ldots, d$, then as N and $T \to \infty$

$$|\hat{f}^{(1)}(x) - f(x)| = \mathcal{O}_p(T^{-\frac{2}{d+4}} log(N) + N^{-1}) \tag{2.22}$$

(b) If the linear model (2.16) holds, then as N and $T \to \infty$

$$|\hat{f}^{(2)}(x) - f(x)| = \mathcal{O}_p(T^{-\frac{1}{2}} log(N) + N^{-1}) \tag{2.23}$$

This theorem deserves some comments. As we expected, the rates of convergence depend on the rates obtained for estimating the individual functions h_i, where only T, the number of observations per firm, was involved, but here a factor of $log(N)$ appears. So T must be large as above but still has to be larger than $log(N)$: $T^{-\frac{2}{d+4}} log(N)$ must be small in the nonparametric case and even in the linear case, $T^{-\frac{1}{2}} log(N)$ must be small. Further, N itself has to be large in order to provide reasonable estimates of the frontier but here the order is N^{-1}.

In conclusion, for the general approach of this section, when individual functions h_i are of interest, we need a large value of T and if the frontier itself is of interest, we need large values of both N and T. Note that this is true for the nonparametric estimator and for the individual linear models as well (except that in the latter we avoid the curse of dimensionality).

In order to improve the rate of convergence of the estimators obtained for $h_i(x)$ we need to introduce some structure or restrictions on the model. Section 3 will propose such a restriction by considering firm effect models.

Remark 1 Stochastic version of DEA and FDH. *The above estimators are reasonable if and only if the basic assumptions of the general setup hold. In particular, assumption (iv) implies some homogeneity between the inputs x_{it} : the densities ρ_{it} are individual ones but the support D is common. This may be unrealistic in some applications (if for instance, the different firms work at two different levels). In that case, the estimation of the best practice frontier by $\hat{f}^{(1)}$ or $\hat{f}^{(2)}$ would be obtained, for some x, by unreasonable extrapolation.*

An alternative approach might be to use the nonparametric estimator above $\hat{h_i}^{(1)}$ *to eliminate, in a nonparametric consistent way, the random noise from* y_{it} *and to use the fitted values of the output* $\hat{h_i}^{(1)}(x_{it})$ *along with the* x_{it} *in a FDH program (or a DEA program if some convexity assumption is added on the attainable production set). This would provide another estimated frontier and the usual procedures could then be used to compute the resulting measures of efficiency.*

The rates of convergence of these estimators are still to be investigated and this idea will not be pursued in this paper. However, as a conjecture, we think that the rate of convergence could, at the best, be equal to those obtained for $\hat{f}^{(1)}$*.*

This attractive two-step approach can be seen as a stochastic version of the DEA and FDH deterministic frontiers.

Very often, the quantities of economic interest are the **individual inefficiencies** of each firm. If we were interested in efficiency estimation in this general setup, we could of course define the following quantities as the basic ingredients for this purpose. Consider indeed

$$\alpha_{it} = h_i(x_{it}) - h(x_{it}) \tag{2.24}$$

where $h(x)$ is the average production frontier with respect to the density ϕ_x

$$h(x) = E(h_i(x)) \tag{2.25}$$

The quantity α_{it} represents indeed for each observed situation (i,t) the distance between the average production level of the firm i and the average level of production over all the technology. The α's could be estimated by replacing h_i by the obtained estimators, h by some empirical average over i and then, some inference could be conducted on particular structural forms of dependence on i and/or on t.

Of course, the statistical performance of the obtained inference will be relatively poor unless T and N are both large enough. Due to that, we will not pursue this general approach, but, in section 4, a global measure of efficiency for each firm will be analysed by averaging the α_{it} over time and using some general nonparametric estimator of h. It will turn out that the rates of convergence are better than those obtained above.

3 A firm effect model.

In this approach we suppose that each firm shares the same technological production function, but that the differences between them are captured by a location effect α_i. This is the most basic approach proposed in the frontier literature with panel data (see e.g. Schmidt and Sickles(1984)). In this case, we additionally rely on the following qualitative model assumption

$$h_i(x) = h(x) + \alpha_i \tag{3.1}$$

where h, defined as before, represents a general production function, common to all firms, and $\alpha_i \in \mathcal{R}$ represents the individual effect of the i^{th} firm. The identifiability condition on the α's corresponds to the fact that h is the mean of the h_i

$$E(\alpha_i) = 0. \tag{3.2}$$

This model could be called a "semi-parametric" model since it is composed of one nonparametric part (the common unspecified production function h) and a parametric part, the efficiency terms modeled by the additive $\alpha_i \in \mathcal{R}, i, \ldots, N$.

3.1 Estimation of h

For the estimation of the "common" production function, we may adopt the same approach as we did in the more general setup.

We can propose a nonparametric estimator of the Nadaraya-Watson type using all the NT observations and obtain

$$\hat{h}^{(1)}(x) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} K(b^{-1} \star (x - x_{it}))y_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} K(b^{-1} \star (x - x_{it}))}. \tag{3.3}$$

A parametric estimation of h is based on the additional model assumption

$$h(x) = X'\beta + \beta_0 \tag{3.4}$$

where $\beta \in \mathcal{R}^d$ and $\beta_0 \in \mathcal{R}$. When using least squares to determine estimates $\hat{\beta}$ and $\hat{\beta}_0$, we obtain the following estimator of h

$$\hat{h}^{(2)}(x) = X'\hat{\beta} + \hat{\beta}_0. \tag{3.5}$$

The basic difference between this firm effect approach and the general model is that now we can use all the NT observations to estimate **one** unknown function. The model is of course more restrictive but the increase in accuracy is particularly important when using the fully nonparametric estimator.

It is straigthforward to show that indeed, under the same regularity assumptions, the rates of convergence are the same as those given in Theorem 1 where T is replaced by NT. However, due to the fact that in the estimation of h an implicit average effect over the N firms is present, an additional error of order $\mathcal{O}_p(N^{-\frac{1}{2}})$ is introduced.

As a consequence, the error of estimation for the nonparametric estimator $\hat{h}^{(1)}$ is thus of order $\mathcal{O}_p((NT)^{-\frac{2}{d+4}} + N^{-\frac{1}{2}})$. The standard rate for the fully parametric model is $\mathcal{O}_p((NT)^{-\frac{1}{2}} + N^{-\frac{1}{2}})$ where the additional error of order $\mathcal{O}_p(N^{-\frac{1}{2}})$ is again due to the normalization $E(\alpha_i) = 0$.

Nevertheless, the "curse of dimensionality" problem does not disappear but is less stringent since there, T is replaced by NT. However, even with NT observations it will usually not make sense to apply the estimator $\hat{h}^{(1)}$ if d is too large

relative to NT [5].

Remark 2 *As an alternative nonparametric estimator of h, one could also think of defining an explicit empirical average over $i = 1, \ldots, N$ of the nonparametric estimators of h_i. The following estimator*

$$\hat{h}^{(3)}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} K(b^{-1} \star (x - x_{it}))y_{it}}{\sum_{t=1}^{T} K(b^{-1} \star (x - x_{it}))}.$$

may indeed be viewed as an estimator of h defined by the empirical counterpart of $E(h_i)$

$$h(x) = \frac{1}{N} \sum_{i=1}^{N} h_i(x).$$

As a conjecture, the rate of convergence of this estimator is given by $\mathcal{O}_p((NT)^{-\frac{2}{d+4}} + T^{-\frac{2}{d}})$ for a suitable choice of the bandwith b. If T is large relative to N and d is not too large, the latter estimator becomes a competitor to the preceding one.

3.2 Estimation of α_i and of the frontier

Having determined an estimator $\hat{h}$ of h, estimators $\hat{\alpha}_i$ of each individual parameter α_i are obtained by least squares :

$$\sum_{t=1}^{T} (y_{it} - \hat{h}(x_{it}) - \hat{\alpha}_i)^2 = \min_{\alpha \in \mathcal{R}} \sum_{t=1}^{T} (y_{it} - \hat{h}(x_{it}) - \alpha)^2. \tag{3.6}$$

The solution is immediate

$$\hat{\alpha_i} = \frac{1}{T} \sum_{t=1}^{T} (y_{it} - \hat{h}(x_{it})). \tag{3.7}$$

where $\hat{h}$ is one of the estimators of h proposed above. The estimates of h and of the α's now provide the following estimators of the individual functions h_i, for $i = 1, \ldots, N$

$$\hat{h}_i^{(3)}(x) = \hat{h}^{(1)}(x) + \hat{\alpha}_i^{(1)} \tag{3.8}$$

$$\hat{h}_i^{(4)}(x) = \hat{h}^{(2)}(x) + \hat{\alpha}_i^{(2)} \tag{3.9}$$

[5] It sould be noted that here again, the problem of dimensionality of x could be overcome by looking to other nonparametric techniques like the generalized additive model. In our case here the additional model assumption might be written

$$h(x) = h_1(x_1) + h_2(x_2) + \ldots + h_d(x_d).$$

Similarly, the estimates of the frontier function are provided by

$$\hat{f}^{(3)}(x) = \max_{i=1,\ldots,N} \hat{h}_i^{(3)}(x) = \hat{h}^{(1)}(x) + \max_{i=1,\ldots,N} \hat{\alpha}_i^{(1)} \tag{3.10}$$

$$\hat{f}^{(4)}(x) = \max_{i=1,\ldots,N} \hat{h}_i^{(4)}(x) = \hat{h}^{(2)}(x) + \max_{i=1,\ldots,N} \hat{\alpha}_i^{(2)} \tag{3.11}$$

The speed of convergence of the obtained estimators are given by the two next Theorems.

Theorem 3 *Assume the firm effect model (3.1) holds in addition to the conditions of Theorem 1, we obtain for every x in the interior of D :*

(a) If $b_j \sim (NT)^{-\frac{1}{d+4}}$, for $j = 1, \ldots, d$, then as N and $T \to \infty$

$$|\hat{h_i}^{(3)}(x) - h_i(x)| = \mathcal{O}_p((NT)^{-\frac{2}{d+4}} + T^{-\frac{1}{2}} + N^{-\frac{1}{2}}) \tag{3.12}$$

(b) If the linear firm effect model (3.4) holds, then as N and $T \to \infty$

$$|\hat{h_i}^{(4)}(x) - h_i(x)| = \mathcal{O}_p((NT)^{-\frac{1}{2}} + T^{-\frac{1}{2}}) \tag{3.13}$$

The reader can appreciate the differences with Theorem 1, in particular for the nonparametric estimators $\hat{h}_i^{(1)}$ and $\hat{h}_i^{(3)}$. For the frontier we obtain :

Theorem 4 *Under the conditions of Theorem 3, we obtain for every x in the interior of D :*

(a) If $b_j \sim (NT)^{-\frac{1}{d+4}}$, for $j = 1, \ldots, d$, then as N and $T \to \infty$

$$|\hat{f}^{(3)}(x) - f(x)| = \mathcal{O}_p(((NT)^{-\frac{2}{d+4}} + T^{-\frac{1}{2}} + N^{-\frac{1}{2}})log(N)) \tag{3.14}$$

(b) If the linear firm effect model (3.4) holds, then as N and $T \to \infty$

$$|\hat{f}^{(4)}(x) - f(x)| = \mathcal{O}_p(T^{-\frac{1}{2}}log(N) + N^{-1}) \tag{3.15}$$

Note again the difference with the Theorem 2 for the nonparametric estimators $\hat{f}^{(1)}$ and $\hat{f}^{(3)}$. The results for the parametric estimators $\hat{h_i}^{(4)}$ and $\hat{f}^{(4)}$ have already be shown in Park and Simar (1994).

Remark 3 Comparison with DEA and FDH rates of convergence. *It would be interesting to compare the rates of convergence of our nonparametric estimators of the frontier $\hat{f}^{(1)}$ in the general case (theorem 2), and $\hat{f}^{(3)}$ in the firm effect model (theorem 4) with the rates of convergence of DEA or FDH type estimators. This comparison is not easy since the models are quite different (DEA and FDH are deterministic and do not allow for noise); further, the rates of convergence obtained by Korostelev, Simar and Tsybakov(1995a) and (1995b) do not*

take into account the panel structure of the data and finally, their rates of convergence are expressed in terms of distance measures between the attainable sets of production rather than on the values of the production frontier itself. However, in our presentation, FDH and DEA models could be written along our notations as in (2.1) :

$$y_{it} = h(x_{it})$$

i.e. no noise is allowed ; the frontier $f(x)$ can be defined (in the case of only one onput y) as above (see (2.2)), additional assumptions are imposed on f, concavity and monotonicity for DEA and monotonicity for FDH. Note that this is a very restrictive model in the presence of panel data (compare with our models (2.1) and (3.1)).

To summarize, the rates of convergence for DEA and FDH are in this case of order $\mathcal{O}_p((NT)^{-\frac{2}{2+d}})$ and $\mathcal{O}_p((NT)^{-\frac{1}{1+d}})$ respectively when the Lebesgue measure of the symmetric difference between the true attainable production set and its estimated conterpart serves as the criterion for measuring the discrepancy between the unknown production process and its estimator.

We see that here again the "curse of dimensionality", as in most of the nonparametric statistical techniques, provides low rates of convergence for large values of d. Since the DEA and FDH models are much more restrictive, the rates are slightly better than in our stochastic framework here (compare the rates of theorems 2 and 4, the presence of the factor $log(N)$) but those rates are of the same order of magnitude for large values of T and N.

So the apparent low rates of convergence obtained here for stochastic frontier estimators are shared with those obtained for the deterministic popular DEA and FDH techniques. That is the price to pay for nonparametric techniques.

4 Measures of efficiencies.

Under the firm effect model (3.1) the estimator of $\alpha_i^{(1)}$ provided by least squares is given by

$$\hat{\alpha_i}^{(1)} = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{h}^{(1)}(x_{it})). \tag{4.1}$$

As a simple corollary of Theorem 4, we obtain

$$|\hat{\alpha_i}^{(1)} - \alpha_i| = \mathcal{O}_p((NT)^{-\frac{2}{d+4}} + T^{-\frac{1}{2}} + N^{-\frac{1}{2}}) \tag{4.2}$$

This estimator is the simplest estimator of the individual firm effects under the firm effect model. However it is not the most efficient.

Further, in this section we would like to approach the problem of estimation of efficiencies in the more general setup of section 2 , where the α_{it} were defined by (2.24).

Suppose that instead of the general model for efficiencies discussed in section 2, we consider an "average over time" measure of efficiency of each firm

$$\alpha_i = \frac{1}{T}\sum_{t=1}^{T}\alpha_{it} = \frac{1}{T}\sum_{t=1}^{T}h_i(x_{it}) - \frac{1}{T}\sum_{t=1}^{T}h(x_{it}) \tag{4.3}$$

Note that this does not imply the firm effect model (3.1) for the individual production functions but can be defined in the more general formulation of section 2 where only model (2.1) is required. Of course, the definition of α_i here is coherent and implied by the firm effect model.

Consider now, for $i = 1, \ldots, N$, the following natural estimator $\hat{\alpha}_i^*$ of α_i, obtained by solving the following least squares problem

$$\sum_{t=1}^{T}(y_{it} - \hat{h}^{(1)}(x_{it}) - \hat{\alpha}_i^*)^2 = \min_{\alpha\in\mathcal{R}}\sum_{t=1}^{T}(y_{it} - \hat{h}^{(1)}(x_{it}) - \alpha)^2. \tag{4.4}$$

The solution is immediate

$$\hat{\alpha_i^*} = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{h}^{(1)}(x_{it})), \tag{4.5}$$

Formally, the two estimators $\hat{\alpha_i^*}$ and $\hat{\alpha_i}^{(1)}$ are identical but, as pointed in Theorem 5, the **bandwith** used for estimating $\hat{h}^{(1)}$ will be different. Indeed, the choice $b_j \sim (NT)^{-\frac{1}{d+4}}, j = 1, \ldots, d$ represents the optimal bandwidth for estimating $h(x)$. It balances bias and variance of the estimator. However estimation of α_i involves a further averaging over the individual values $y_{it} - \hat{h}^{(1)}(x_{it}), t = 1, \ldots, T$. This averaging reduces the variance of the estimators $\hat{\alpha_i^*}$ compared to the estimator $\hat{h}^{(1)}(x_{it})$ at specific points. Hence, for estimating the α_i, it is preferable to use undersmoothing bandwidth $b_j \sim (NT)^{-\frac{1}{d+2}} = o((NT)^{-\frac{1}{d+4}})$ which provides a further bias reduction. The following Theorem summarizes the properties of $\hat{\alpha_i^*}$.

Theorem 5 *Assume that the conditions of Theorem 1 hold. Then as N and $T \to \infty$, if $b_j \sim (NT)^{-\frac{1}{d+2}}$, for $j = 1, \ldots, d$, we obtain*

(a)

$$|\hat{\alpha}_i^* - \alpha_i| = \mathcal{O}_p(T^{-\frac{1}{2}} + N^{-\frac{1}{2}} + (NT)^{-\frac{2}{d+2}}) \tag{4.6}$$

(b)

$$|\max_{i=1,\ldots,N}\hat{\alpha}_i^* - \max_{i=1,\ldots,N}\alpha_i| = \mathcal{O}_p((T^{-\frac{1}{2}} + N^{-\frac{1}{2}} + (NT)^{-\frac{2}{d+2}})log(N)) \tag{4.7}$$

Note that part (b) of the Theorem gives the asymptotic properties for the estimator of the frontier level itself ; it shows the price to pay when using a nonparametric frontier model when compared with the rate convergence obtained

in Park and Simar (1994) where a parametric firm effect model was analysed (the presence of the term of order $\mathcal{O}_p((NT)^{-\frac{2}{d+2}})$).

Note also that the result is not so dramatic : if T is of the same order as N, and if $d \leq 6$, the error of estimation for the α_i is of the standard order $\mathcal{O}_p(T^{-\frac{1}{2}})$.

Remark 4 *The definition (4.3) could also be used to provide estimators of α_i by replacing $h_i(x)$ and $h(x)$ by their nonparametric estimators $\hat{h_i}^{(1)}(x)$ and $\hat{h}^{(1)}(x)$ respectively. The resulting estimator is consistent but it can be shown that the rate of convergence is lower than for $\hat{\alpha}_i^*$. The main reason is that the two estimators $\hat{h_i}^{(1)}$ and $\hat{h}^{(1)}$ have a different bias which does not simplify by the difference in (4.3) (however, the variance of the estimator is not a problem here due to the averaging over T). To overcome this bias effect, the baseline bandwidth would be so small that the number of observations necessary to obtain sensible estimators would be too large.*

Remark 5 *Note that in the particular case of x_{it} being i.i.d. d-variate random variables, $\rho_{11} = \rho_{12} = \ldots = \rho_{NT} = \rho$, α_i might be estimated through*

$$\hat{\alpha_i}^{(2)} = \frac{1}{T}\sum_{t=1}^{T} y_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}. \tag{4.8}$$

It can easily be derived that $\hat{\alpha_i}^{(2)}$ has a rate of convergence of order $\mathcal{O}_p(T^{-\frac{1}{2}})$ independent of the dimension d.

5 Empirical illustration.

In order to illustrate the above results, we consider the analysis of labor efficiencies of 17 railway companies observed over a period of 14 years ($T = 14, N = 17$). The purpose here is just to illustrate numerically the proposed estimators. The numbers of observations N and T are too small to draw definite conclusions about the obtained measures of efficiency of the railways, but allow to show from a practical point of view how the procedures work.

Data on the activity of railways can be found in the annual reports of the Union Internationale des Chemins de Fer (U.I.C.). The railways retained for the illustration are given in Table 1.

Insert here Table 1

The data are available for each network on an annual basis. In this study, we use the period 1970-1983. This provides 238 observations on the whole set of variables. We choose in our example the following set of variables :

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

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

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

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

All the variables are in logarithms and have been adjusted for time trend effect. Note that the division of labor by the total length of the network allows to eliminate a size effect in the input y. The variable x_1 represents a rough measure of the output (the demand) of the railways and is also corrected for the size effect, whereas x_2 characterizes some aspects of the demand and x_3 is a physical measure of the density of a network.

Insert here Figure 1

Figure 1 shows the marginal distributions of each of these variables across all the panel of data, the horizontal axes being the 238 observations ranked by networks (according to the order of Table 1). This allows, for instance, to appreciate the sampling variations of the variables (note e.g. the low level of labor/km of Norway (#11) and Sweden (#14) and the particularly high level of Japan (#9)).

Because we analyse a labor function, the supremum in the definition (2.2) of the frontier has to be replaced by an infimum. Accordingly, in the corresponding estimators the max is replaced by a min.

We first provide the results for the estimation of the parametric linear model

$$y_{it} = \beta_0 + X_{it}\beta + \alpha_i + \epsilon_{it} \qquad \text{where } X_{it} = (x_{1it}, x_{2it}, x_{3it}) \tag{5.1}$$

Figure 2 shows the quality of the fit (compare y_{it} with $\hat{h}^{(2)}(x_{it})$) and the sampling distribution of the residuals $\alpha_i + \epsilon_{it}$ obtained from the least-squares estimation of β. The spikes in the latter being interpretable as spikes in the inefficiency distribution.

Insert here Figure 2

For each network, the distances between y_{it} and $\hat{h}^{(2)}(x_{it})$ provide, by (3.7), the estimators of α_i. Table 2 gives the estimates of the parameters β of the linear model and the estimates $\hat{\alpha_i}^{(2)}$ for each network, ranked in decreasing order of labor efficiencies.

Insert here Table 2

Note that all the coefficients of the linear model for the common production function $h(x)$ have the right sign and are significant. The estimated values of α_i and the ranking of the networks are of course relying on the restrictive linear hypothesis (3.4). The question now is, what happens with a less restrictive model where a nonparametric h is estimated ?

The same analysis is performed with the Nadaraya-Watson estimator $\hat{h}^{(1)}$ proposed in (3.3). The baseline bandwith b was chosen by cross-validation. In order to be more robust relative to the some eventual time dependence of the observations within a network, we prefered to use a leave-one network-out approach to estimate the cross-validation criterion. The CV criterion (2.14) particularizes here as follows :

$$\sum_{i,t} (y_{it} - \hat{h}^{(1)}_{-i}(x_{it}))^2 \tag{5.2}$$

where $\hat{h}_{-i}^{(1)}$ is $\hat{h}^{(1)}$ computed by leaving out all the T observations (y_{it}, x_{it}) of the i^{th} network. As shown in Figure 3, the optimal bandwith for the estimation of h is around 1.86. The second part of Figure 3 deserves some comments : it shows how the mean number of points entering in the computation of $\hat{h}^{(1)}(x)$, for a fixed point x, depends on the baseline bandwidth b. Very soon ($b > 3$) almost all the 238 points are used in the weighted averaging procedure of the kernel estimates (when $b \to \infty, \forall x, \hat{h}^{(1)}(x) \to \bar{y}$). The clustered structure of the data (compare Figure 1) would make unrealistic a Nadaraya-Watson type estimator with too large bandwidths.

Insert here Figure 3

The results of the estimation with the optimal bandwidth $b = 1.86$ are presented in Figure 4 and Table 3. Figure 4 when compared with Figure 2 allows to appreciate the difference of the two estimators of h (the restricted linear one and the unrestricted nonparametric one). The unrestricted estimator allows to follow smoothly the observed y. The distribution of $\alpha_i + \epsilon_{it}$ is smoother than in the linear case : the linear model introduced spurious spikes in the distribution and consequently, spurious inefficiencies. Note that the middle of Figure 4 shows the number of observations used to estimate $h(x)$ for each x_{it}.

The resulting efficiency scores $\hat{\alpha_i}^{(1)}$ (computed from (4.1)) are provided by Table 3 : note the differences of the obtained estimates and the differences between the ranking with the linear case (compare Table 2) for a majority of networks. If this would have been obtained with a larger number of observations, we could conclude that the restrictive linear model is incorrect.

Insert here Figure 4

Insert here Table 3

As pointed in section 4, if the main interest is the estimation of the efficiencies α_i, an undersmooth estimator of h is preferable (see Theorem 5). We propose here to use a baseline bandwith of $b = 1.7$. The results are provided by Figure 5 and Table 4 : the results are very similar to the preceding case.

Insert here Figure 5

Insert here Table 4

Finally, Figure 6 shows the role of the bandwidth in the smoothing procedure of the Kernel estimator. Clearly, as b tends to zero, the estimator becomes an interpolation and as b increases, the estimator oversmooths the data.

Insert here Figure 6

6 Conclusions.

We have shown in the paper that it is possible to estimate nonparametric stochastic frontier models when a panel of data is available. As in any nonparametric model, the curse of dimensionality imposes a large number (T and N) of observations in order to get reasonable estimates of the frontier : in all the analyzed models, the rates of convergence are provided. The situation is more favorable when one is only interested to obtain estimates of the (time-averaged) relative efficiency of the

different firms.

This is a first paper in this framework : the nonparametric stochastic frontier models have been presented in a general framework with some simple estimators and the statistical properties have been investigated under some familiar regularity conditions which facilitate the presentation. No doubt that in further works, some of these conditions (like homoscedasticity, time independence ...) will be relaxed and that more sophisticated models and techniques will be developed.

In our illustration with railway companies, despite the small values of N and T, we have shown the problems induced by using a retrictive linear model when it could be incorrect. This might provide spurious inefficiencies and wrong efficiency scores which penalizes some networks.

References

[1] Aigner, D.J. and S.F. Chu (1968), On estimating the industry production function, *American Economic Review*, 58, 826-839.

[2] Aigner, D.J., Lovell, C.A.K. and P. Schmidt (1977), Formulation and estimation of stochastic frontier models, *Journal of Econometrics*, 6, 21-37.

[3] Andersen, P. and N.C. Petersen (1993), A procedure for ranking efficient units in data envelopment analysis, *Management Science*, 39, 1261-1264.

[4] Banker, R.D. (1993), Maximum Likelihood, consistency and data envelopment analysis : a statistical foundation, *Management Science*, 39,10,1265-1273.

[5] Battese, G.E. and T.J. Coelli (1992), Fontier production functions, technical efficiency and panel data, *Journal of Productivity Analysis*, 3, 153-169.

[6] Charnes, A., Cooper W.W. and E. Rhodes (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2, 429-444.

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

[8] Craven, P. and G. Wabha (1979), Smoothing noisy data with spline functions : estimating the correct degree of smoothing by generalized cross-validation, *Numer. Math.*, 31, 377-403.

[9] Deprins, D., Simar, L. and H. Tulkens (1984), Measuring labor inefficiency in post offices, in *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243-267.

[10] Färe, R., Grosskopf, S. and C.A.K. Lovell (1985), *The Measurement of Efficiency of Production*. Boston, Kluwer-Nijhoff Publishing.

[11] Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society*, A(120), 253-281.

[12] Greene, W.H. (1980), Maximum likelihood estimation of econometric frontier, *Journal of Econometrics*, 13, 27-56.

[13] Hall, P., Härdle, W. and L. Simar (1991), Iterated bootstrap with application to frontier models, CORE Discussion paper 9121, UCL, to appear in the *Journal of Productivity Analysis*.

[14] Härdle, W. (1990),*Applied Nonparametric Regression*.Cambridge University Press, Cambridge.

[15] Gasser, T. , Kneip, A. and W. Köhler (1991), A flexible and fast method for automatic smoothing, *Journal of the American Statistical Association*,86,643-652.

[16] Gasser, T. , Müller, H.G. and V. Mammitzsch (1985), Kernels for nonparametric curve estimation, *Journal of the Royal Statistical Society B*, 47,238-252.

[17] Gasser, T., Stoka, L. and C. Jennen-Steinwetz (1986), Residual variance and residual pattern in nonlinear regression, *Biometrika*, 73, 623-633.

[18] Hall, P., Kay, J.W. and D.M. Titterington (1990), Asymptotically optimal difference based estimation of variance in nonparametric regression, *Biometrika*, 77,521-529.

[19] Hastie, T. and R.J. Tibshirani (1990), *Generalized Additive Models*. Chapman and Hall, London.

[20] Jondrow, J., Lovell, C.A.K., Materov, I.S. and P. Schmidt (1982), On the estimation of technical inefficiency in stochastic frontier production models, *Journal of Econometrics*, 19, 233-238.

[21] Kneip, A. and L. Simar (1995) , Stochastic nonparametric frontier estimation with panel data, forthcoming manuscript.

[22] Korostelev, A., Simar, L. and A. Tsybakov (1995a), Efficient estimation of monotone boundaries, *The Annals of Statistics*, 23(2), 476-489.

[23] Korostelev, A., Simar, L. and A. Tsybakov (1995b), On estimation of monotone and convex boundaries, *Pub. Inst. Stat. Univ. Paris*, XXXIX, 1, 3-18.

[24] Kumbhakar, S.C. (1990), Production frontier, panel data and time-varying technical efficiency, *Journal of Econometrics*, 46, 201-212.

[25] Lee, Y.H. and P. Schmidt (1993), A production frontier model with flexible temporal variation in technical efficiency, in *The Measurement of Productive Efficiency Techniques and Applications*, eds. Fried, H, Lovell, C.A.K. and S.S. Schmidt, Ch. 8, 237-255, Oxford University Press.

[26] Lovell, C. A. K., Walters, L.C. and L.L. Wood (1995), Stratified Models of Education Production Using DEA and Regression Analysis, in *Data Envelopment Analysis: Theory, Methods, and Applications*, eds. A. Charnes, W. W. Cooper, A. Y. Lewin, and L. M. Seiford, Boston: Kluwer Academic Publishers.

[27] Müller, H.G. (1988), *Nonparametric regression analysis of longitudinal data*, Springer Verlag, Heidelberg.

[28] Park, B. and L. Simar (1994), Efficient semiparametric estimation in a stochastic frontier model, *Journal of the American Statistical Association*,89(427),929-936.

[29] Park, B., Sickles, R.C. and L. Simar (1993), Stochastic panel frontiers, a semiparametric approach, Discussion paper 9303, Institut de Statistique, UCL, Louvain-la-Neuve, Belgium.

[30] Ruppert, D. and M. Wand (1994), Multivariate locally weighted least squares regression, *Annals of Statistics*, to appear.

[31] Schmidt, P, and R.C. Sickles (1984), Production frontier and panel data, *Journal of Business and Economic Statistics*,3, 171-203.

[32] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[33] Simar, L. (1992), Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping, *Journal of Productivity Analysis*, 3, 167-203.

[34] Simar, L. and P. Wilson (1995), Sensitivity analysis of efficiency scores : how to bootstrap in nonparametric frontier models, Discussion paper, Institut de Statistique, UCL, Louvain-la-Neuve, Belgium.

[35] Stone, C.J. (1982), Optimal global rates of convergence for nonparametric regression, *Annals of Statistics*, 14, 590-606.

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

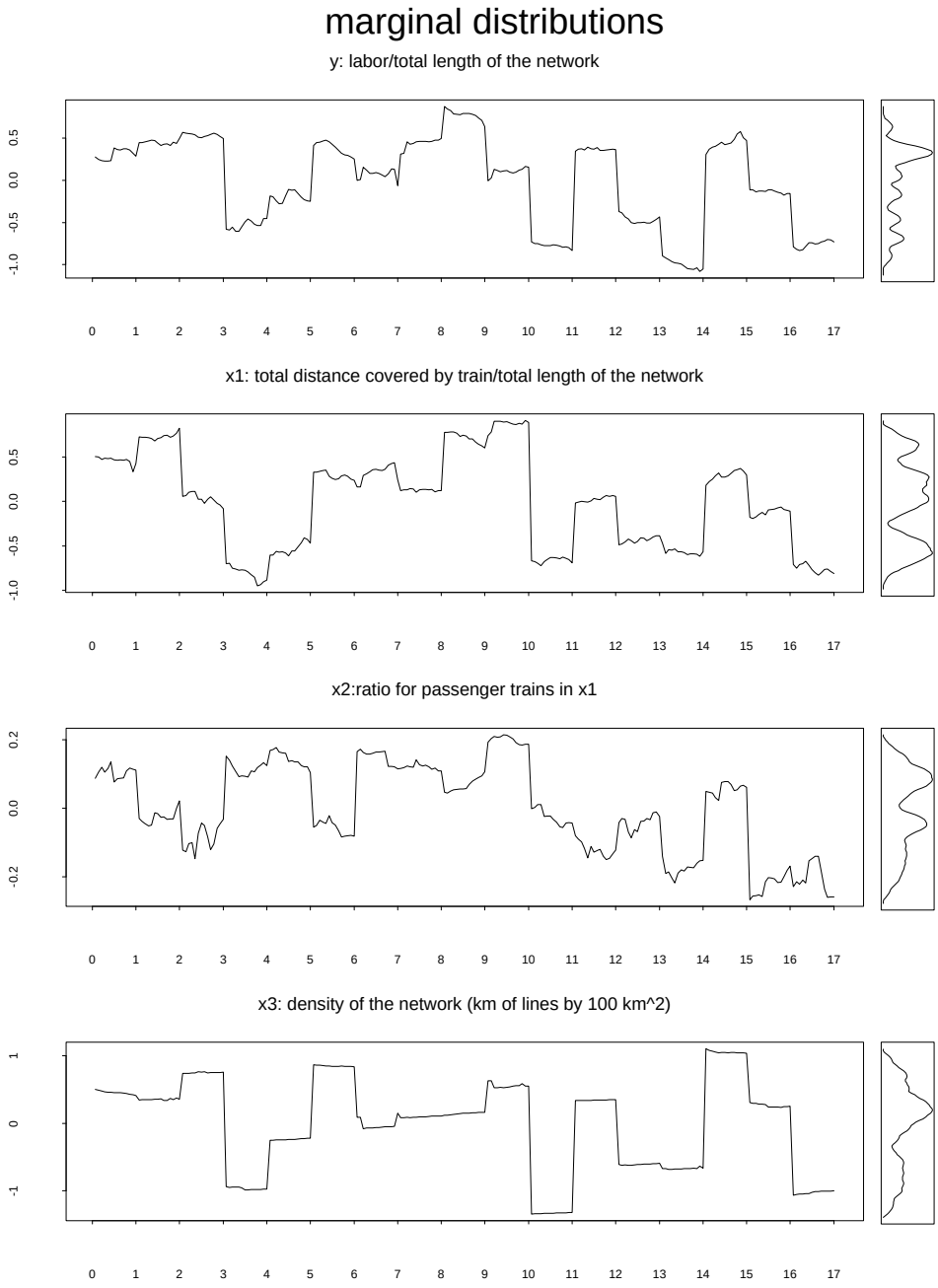
marginal distributions
y: labor/total length of the network
x1: total distance covered by train/total length of the network
x2:ratio for passenger trains in x1
x3: density of the network (km of lines by 100 km^2)

Figure 1

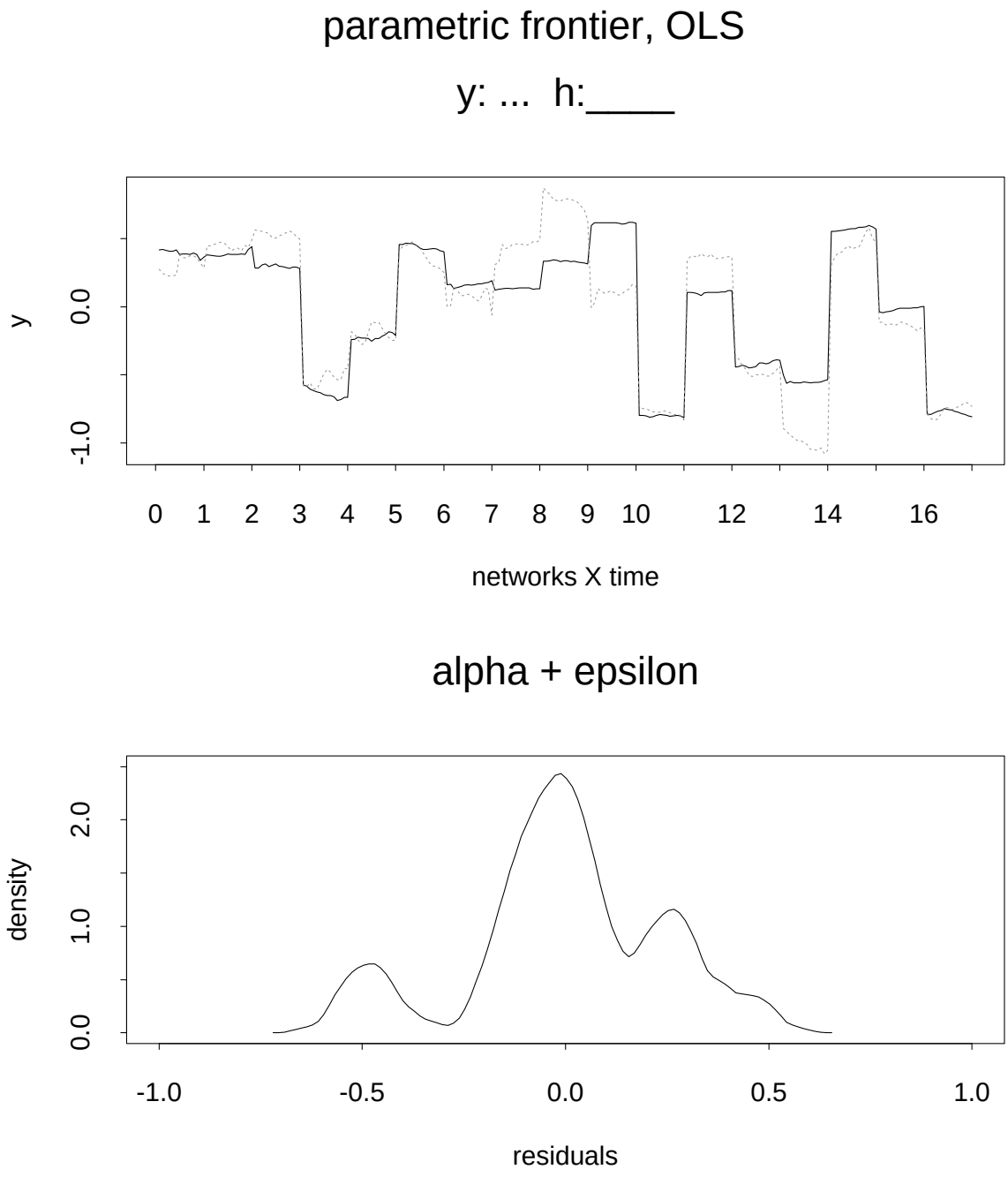
parametric frontier, OLS
y: ... h:____
y
-1.0
0.0
0 1 2 3 4 5 6 7 8 9 10 12 14 16
networks X time
alpha + epsilon
density
0.0
1.0
2.0
-1.0 -0.5 0.0 0.5 1.0
residuals

Figure 2

bandwidth choice
cross validation
cv
0.1
0.2
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0
bandwidth
mean number of obs. in (x - bw,x + bw)
number of obs.
0
100
200
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0
bandwidth

Figure 3

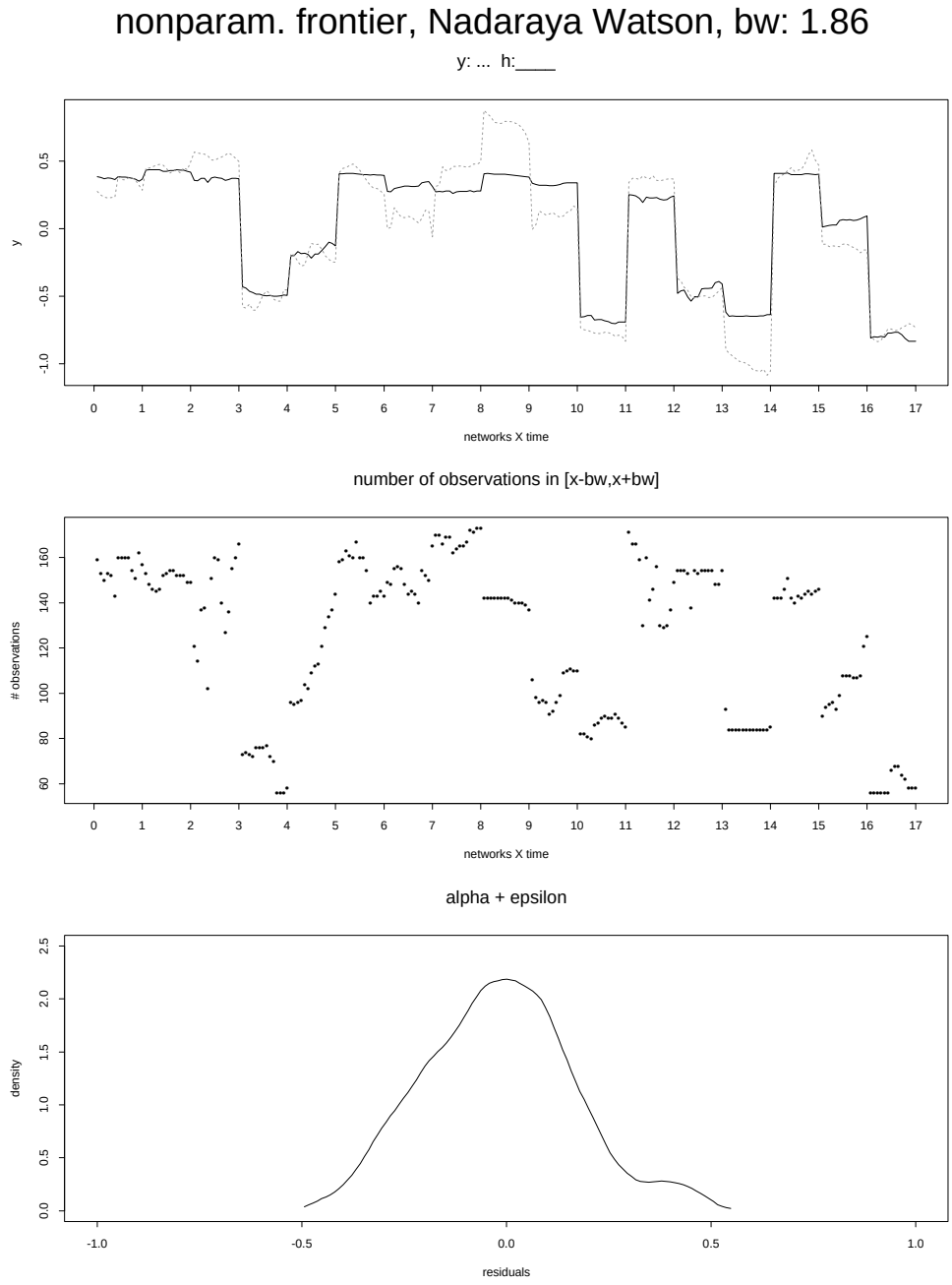
nonparam. frontier, Nadaraya Watson, bw: 1.86
y: ... h:____
y
networks X time
number of observations in [x-bw,x+bw]
observations
networks X time
alpha + epsilon
density
residuals

Figure 4

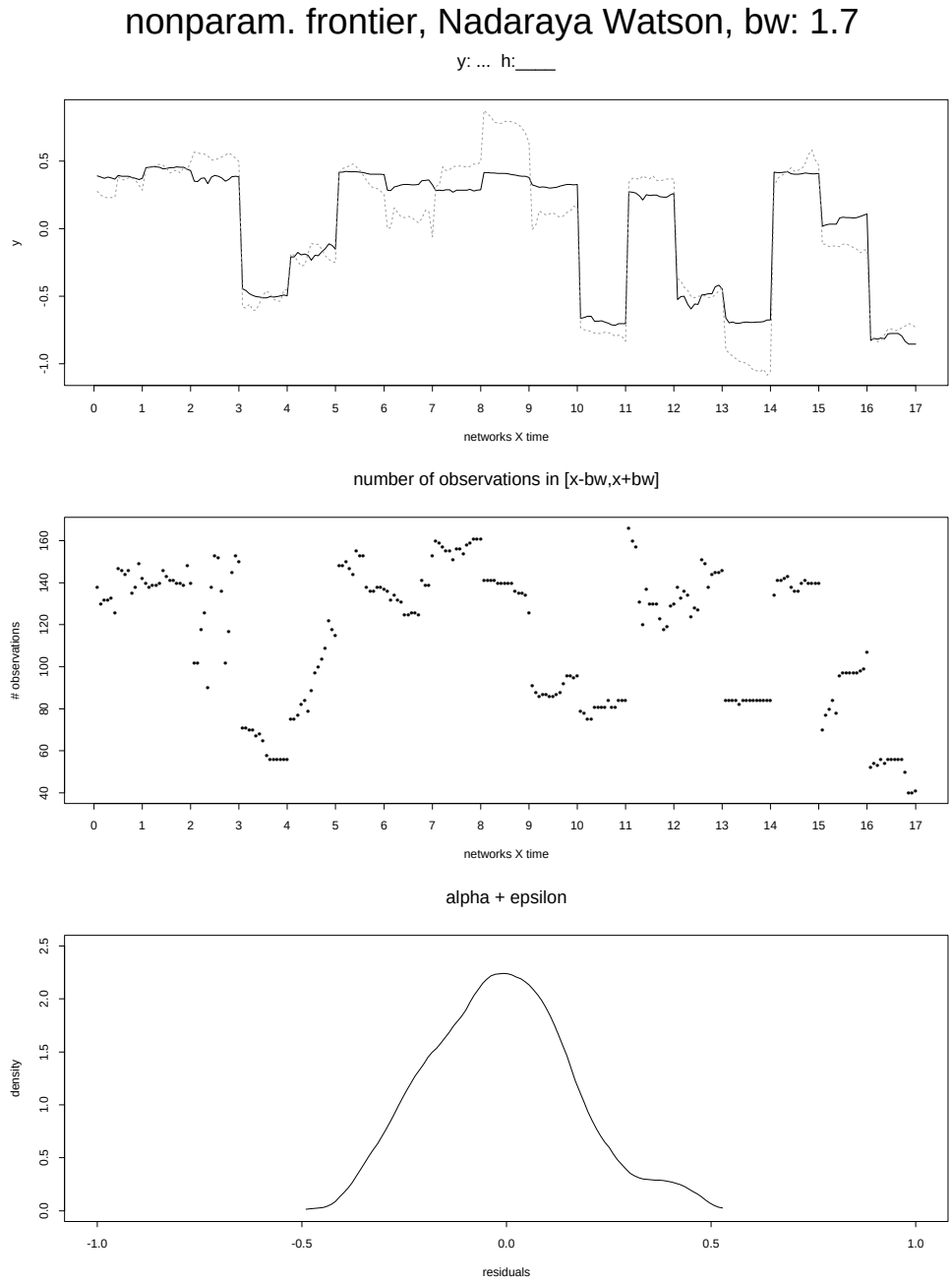
nonparam. frontier, Nadaraya Watson, bw: 1.7
y: ... h:____
y
networks X time
number of observations in [x-bw,x+bw]
observations
networks X time
alpha + epsilon
density
residuals

Figure 5

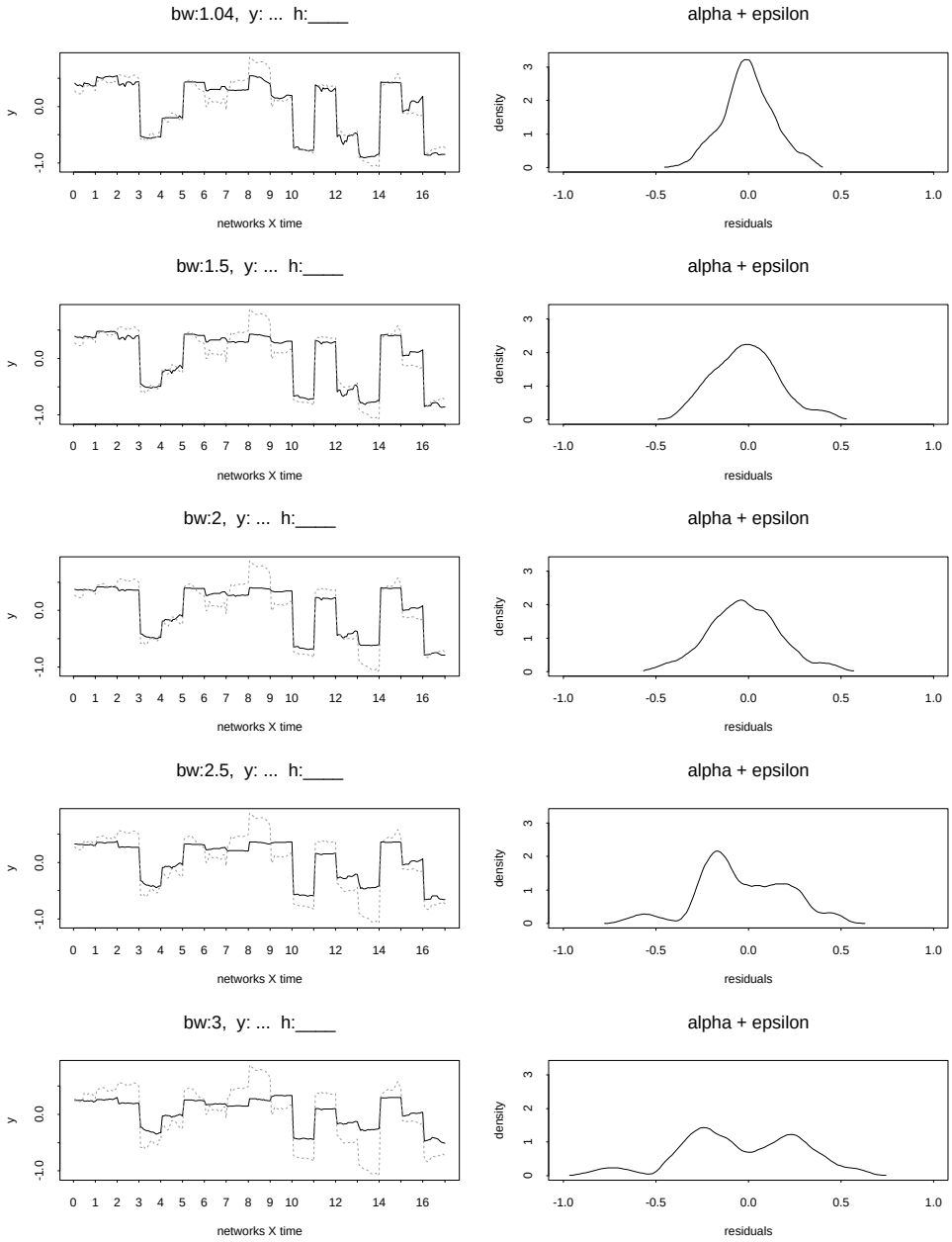
nonparm. frontier, Nadaraya Watson
bw:1.04, y: ... h:____
alpha + epsilon
bw:1.5, y: ... h:____
alpha + epsilon
bw:2, y: ... h:____
alpha + epsilon
bw:2.5, y: ... h:____
alpha + epsilon
bw:3, y: ... h:____
alpha + epsilon
y
density
networks X time
residuals

Figure 6

No	Network	Country	No	Network	Country
1	BR	Great Britain	10	NS	Netherlands
2	CFF	Switzerland	11	NSB	Norway
3	CFL	Luxemburg	12	ÖBB	Austria
4	CH	Greece	13	RENFE	Spain
5	CP	Portugal	14	SJ	Sweden
6	DB	Germany	15	SNCB	Belgium
7	DSB	Denmark	16	SNCF	France
8	FS	Italy	17	VR	Finland
9	JNR	Japan			

Table 1: Railway Companies

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

Residual Standard Error = 0.2445, Multiple R-Square = 0.7852 N = 238,
F-statistic = 285.1411 on 3 and 234 df, p-value = 0

	coef	std.err	t.stat	p.value
Intercept	0.0000	0.0158	0.0000	1.0000
X1	0.3387	0.0521	6.4973	0.0000
X2	0.4090	0.1347	.0367	0.0027
X3	0.4270	0.0382	11.1735	0.0000

Estimated Efficiencies					
Network	α	shifted α	Network	α	shifted α
NS	-0.512	0.000	CP	0.025	0.537
SJ	-0.452	0.060	NSB	0.028	0.540
SNCB	-0.128	0.384	CFF	0.060	0.572
SNCF	-0.117	0.395	CH	0.110	0.622
DSB	-0.091	0.421	CFL	0.240	0.752
BR	-0.090	0.422	ÖBB	0.262	0.774
DB	-0.058	0.454	FS	0.306	0.818
RENFE	-0.045	0.467	JNR	0.446	0.958
VR	0.018	0.530			

Table 2: OLS estimates

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

frontier estimation with Nadaraya Watson estimator, baseline bandwidth: 1.86
mean number of observations: 124.8319

Estimated Efficiencies					
Network	α	shifted α	Network	α	shifted α
SJ	-0.359	0.000	RENFE	-0.010	0.348
DSB	-0.240	0.119	CFF	0.018	0.377
NS	-0.225	0.134	VR	0.039	0.398
SNCF	-0.187	0.171	SNCB	0.041	0.400
NSB	-0.100	0.259	ÖBB	0.140	0.498
BR	-0.068	0.291	FS	0.166	0.524
CH	-0.048	0.311	CFL	0.169	0.528
CP	-0.030	0.329	JNR	0.382	0.741
DB	-0.021	0.338			

Table 3: nonparametric estimates, baseline bandwidth: 1.86

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

frontier estimation with Nadaraya Watson estimator, baseline bandwidth: 1.7
mean number of observations: 114.4706

Estimated Efficiencies					
Network	α	shifted α	Network	α	shifted α
SJ	-0.314	0.000	CFF	-0.001	0.312
DSB	-0.252	0.062	RENFE	0.034	0.348
NS	-0.212	0.102	SNCB	0.035	0.349
SNCF	-0.201	0.113	VR	0.052	0.366
NSB	-0.089	0.224	ÖBB	0.120	0.433
BR	-0.074	0.240	FS	0.157	0.470
CH	-0.036	0.278	CFL	0.165	0.479
DB	-0.031	0.282	JNR	0.377	0.691
CP	-0.019	0.295			

Table 4: nonparametric estimates, baseline bandwidth: 1.7