Semiparametric Efficient Estimation of AR(1) Panel Data Models

Byeong Park
Department of Statistics
Seoul National University

Robin C. Sickles
Department of Economics
Rice University

Léopold Simar
Institut de Statistique
Université Catholique de Louvain

January 28, 2000

Abstract

This study focuses on the semiparametric efficient estimation of random effect panel models containing AR(1) disturbances. We also consider such estimators when the effects and regressors are correlated (Hausman and Taylor, 1981). One motivation for such a model is the need to estimate a stochastic frontier distance function, isolate the fixed effects estimates, and interpret transformations of them as firm-specific relative efficiencies (Schmidt and Sickles, 1984). In markets in which regulatory constraints have been lessened or done away with, these market shocks may not be adjusted to immediately and may induce a serial correlation pattern in within firm variations. In the banking industry, whose productivity we analyze over the 1980's and 1990's, these aspects of semiparametric efficient estimation become important in order to resolve questions concerning the proper measurement of financial deregulation's impact on productivity. We introduce two semiparametric efficient estimators that make minimal assumptions on the distribution of the random errors, effects, and the regressors and that provide semiparametric efficient estimates of the slope parameters and of the effects. Our estimators extend the previous work of Park and Simar (1995), Park, Sickles, and Simar (1998), and Adams, Berger, and Sickles (1998).

1 Introduction

This study focuses on the semiparametric efficient estimation of random effect panel models containing AR(1) disturbances. We also consider such estimators when the effects and regressors are correlated (Hausman and Taylor, 1981). One motivation for such a model is the need to estimate a stochastic frontier distance function, isolate the fixed effects estimates, and interpret transformations of them as firm-specific relative efficiencies (Schmidt and Sickles, 1984). In markets in which regulatory constraints have been lessened or done away with, these market shocks may not be adjusted to immediately and may induce a serial correlation pattern in within firm variations. In the banking industry, whose productivity we analyze over the 1980's and 1990's, these aspects of semiparametric efficient estimation become important in order to resolve questions concerning the proper measurement of financial deregulation's impact on productivity. We introduce two semiparametric efficient estimators that make minimal assumptions on the distribution of the random errors, effects, and the regressors and that provide semiparametric efficient estimates of the slope parameters and of the effects. Our estimators extend the previous work of Park and Simar (1995), Park, Sickles, and Simar (1998), and Adams, Berger, and Sickles (1998).

Semiparametric efficient estimation has been discussed extensively in the statistics and econometrics literature. Bickel (1982), Newey (1990), Bickel, Klaassen, Ritov, and Wellner (1993) as well as others have developed semiparametric efficient methods and examples. In their article on semiparametric efficient estimation, Park and Simar (1995) introduce a semiparametric efficient estimator for the specific problem of a panel data model, where the distribution of the firm specific heterogeneity is unknown. In the derivation of their estimator, they assume normality of the transitory error as well as independence of the regressors and effects. Park, Sickles, and Simar (1998) extended their model in that they allowed a regressor to be correlated with the effects and explore the impacts of various correlation patterns among effects and regressors on the form of the semiparametric efficient estimator. The statistical assumptions, in particular normality of the transitory error and independence of the effects and regressors, have a direct bearing on the form of the efficient score and information bound (the center pieces of the estimator) and allow the authors to concentrate on the unknown distribution of the effects and also draw similarities to other estimators, such as the within estimator.[1] A change in these assumptions results in a change in the loglikelihood function and nuisance parameter space and, hence, in the semiparametric efficient estimator. Adams, Berger, and Sickles (1998) developed semiparametric efficient

[1] Park, Sickles, and Simar (1998) show that within is efficient when all regressors are correlated with the effects.

estimators using higher dimensioned product kernels for the (log) linear model as well for the semilinear model of Robinson. Others have considered efficient estimation with different assumptions. Chamberlain (1987, 1992) and Arrelano and Bover (1995) discussed efficient estimation with strict exogeneity assumptions. This paper generalizes the semiparametric efficient estimator derived by Park and Simar (1995) and by Park, Simar, and Sickles (1998) by allowing for the distribution of the transitory error to be autocorrelated.

In Section 2 contains our main results: we outline our general panel model and we derive two semiparametric efficient estimators according to the relationships between the regressors and the effects (independence and dependence). Section 3 considers Monte Carlo results. Section 4 outlines the modeling scenario on which our empirical illustration, estimating the efficiency of the U.S. banking industry during its regulatory transition of the 1980's and 1990's, is based. Section 4.3 describes the results from our application in analyzing banking productivity. Section 5 concludes. All the technical proofs are in the appendix.

2 Models and Main Results

The basic model we analyze in this paper is an AR(1) panel model that can be written as:

$$\begin{aligned} Y_{it} &= X_{it}^{\prime}\beta + \alpha_i + \varepsilon_{it} \quad ; \quad i = 1,\ldots,N \; ; \; t = 1,\ldots,T \\ \varepsilon_{it} &= \rho\varepsilon_{i,t-1} + u_{it} \quad ; \quad |\rho| < 1 \end{aligned} \tag{1}$$

where $X_{it} \in \mathbb{R}^d$, $\beta \in \mathbb{R}^d$ and u_{it} are iid random variables from a $N(0,\sigma^2)$. Denoting $X_i = (X_{i1}^{\prime},\ldots,X_{iT}^{\prime})^{\prime}$, (α_i, X_i)'s are iid random variables having unknown density $h(\cdot,\cdot)$ on $\mathbb{R}^{1+dT}$. The support of the marginal density of α is bounded above (or below). This bound B provides the upper level of the production frontier or the lower level of, e.g., the cost frontier. Finally we assume that ε's and (α, X)'s are independent.

Although equation (1) is the generic panel data model, an appealing empirical motivation for the appropriate treatment of (1) is to estimate firm specific efficiency levels in a stochastic panel production frontier model (c.f. Schmidt and Sickles, 1984; Cornwell, Schmidt and Sickles, 1990). There, Y_{it} is the t-th observation on the output of the i-th firm, X_{it} is a vector of the t-th observation of the d inputs of the i-th firm and α_i is an unobservable random effect that captures firm specific inefficiency. The availability of panel data allows identification of realizations of α_i for a particular firm and thus overcomes the limitation of a single cross-section (or time series) which allows only the identification of the expectation of α_i conditional on stochastic noise (Jondrow, Lovell, Materov and Schmidt, 1982). Normality is a rather natural assumption for the within disturbance term based on central limit arguments, while particular parametric distributions for the inefficiency terms are less easily motivated.

We will consider two structures of dependency between X and α: independency in Model 1 and dependency in Model 2. Throughout this section, we consider the time period T fixed.

2.1 Model 1

In this first case, X_i and α_i are supposed to be independent, as with the ε_i. We denote by $h(\cdot)$ the univariate density of the α_i and by $g(\cdot)$ the dT-variate density of X_i and we want to consider the semiparametric efficient estimation of β in Model 1 from the sample $\{(X_i, Y_i) \mid i = 1, \ldots, N\}$ in the presence of the nuisance parameters $(\rho, \sigma^2, h(\cdot), g(\cdot))$.

We will use the basic techniques discussed in Appendix A, with the notations introduced there. Let $Y = (Y_1, \ldots, Y_T)'$, $X = (X_1', \ldots, X_T')$ for the generic of observation (X_i, Y_i) and (α, ε) for the generic of $(\alpha_i, \varepsilon_i)$. Thus, in these notations, (X_t, Y_t) are generics for (X_{it}, Y_{it}), $i = 1, \ldots, N$.

The pdf of (X, Y) can be written as:

$$p(x, y; \beta, \rho, \sigma^2, h, g) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T (1-\rho^2)^{1/2} g(x) \exp\{w^2(\rho, \beta)/(2v^2)\}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}\left[(1-\rho^2)(y_1 - x_1'\beta)^2 + \sum_{t=2}^{T}(y_t - x_t'\beta - \rho(y_{t-1} - x_{t-1}'\beta))^2\right]\right\}$$
$$\times \int \exp\{-(w(\rho, \beta) - u)^2/(2v^2)\} h(u)\, du,$$

where:

$$w(\rho, \beta) = \frac{1}{T(\rho)}\left[(1-\rho^2)(y_1 - x_1'\beta) + (1-\rho)\sum_{t=2}^{T}(y_t - x_t'\beta - \rho(y_{t-1} - x_{t-1}'\beta))\right]$$
$$T(\rho) = (1-\rho^2) + (T-1)(1-\rho)^2$$
$$v^2 = v^2(\rho) = \frac{\sigma^2}{T(\rho)}.$$

It is useful to note that $w(\rho, \beta)$ is a weighted average of $y_1 - x_1'\beta, \ldots, y_T - x_T'\beta$: $w(\rho, \beta) = \sum_{t=1}^{T} c_t(y_t - x_t'\beta)$, where

$$c_t = c_t(\rho) = \begin{cases} (1-\rho)/T(\rho) & ; \text{ for } t = 1 \text{ and } T \\ (1-\rho)^2/T(\rho) & ; \text{ for } t = 2, \ldots, T-1 \end{cases}$$

Note that indeed $\sum_{t=1}^{T} c_t = 1$.

If we define the random variable $W = W(\rho, \beta) = \sum_{t=1}^{T} c_t(Y_t - X_t'\beta)$ we can write:

$$W = \alpha + \sum_{t=1}^{T} c_t \varepsilon_t$$
$$= \alpha + \{(1-\rho^2)\varepsilon_1 + (1-\rho)(u_2 + \ldots + u_T)\}/T(\rho)$$

where the second term follows a $N(0;v^2)$ distribution. Thus the pdf of W is given by:

$$f(w) = f(w;\rho) = \int \frac{1}{v}\phi\left(\frac{w-u}{v}\right)h(u)\,du$$
$$= \frac{1}{\sqrt{2\pi}v}\int \exp\left(-\frac{(w-u)^2}{2v^2}\right)h(u)\,du. \quad (2)$$

Let $Z_t = Z_t(\rho;\beta) = Y_t - X_t'\beta - W$ and denote by $\bar{X}^f$ the weighted average of $X_1,\ldots,X_T$:

$$\bar{X}^f = \bar{X}^f(\rho) = \sum_{t=1}^{\bar{T}} c_t X_t.$$

Let I_f be the Fisher information for location of $h(\cdot)$:

$$I_f = \int \frac{(f^{(1)})^2}{f}(w)\,dw$$

where $f^{(j)}$ denotes the jth derivative of f. Finally defining:

$$\Sigma_1 = E\{(1-\rho^2)(X_1-\bar{X}^f)(X_1-\bar{X}^f)'$$
$$+ \sum_{t=2}^{\bar{T}}(X_t-\bar{X}^f-\rho(X_{t-1}-\bar{X}^f))(X_t-\bar{X}^f-\rho(X_{t-1}-\bar{X}^f))'\};$$
$$\Sigma_2 = E(\bar{X}^f-E\bar{X}^f)(\bar{X}^f-E\bar{X}^f)',$$

we can now write the following theorem:

Theorem 2.1 *The efficient score function and the information bound for estimating β in Model 1 are given by:*

$$\ell_\beta = \frac{1}{\sigma^2}\{(1-\rho^2)Z_1X_1 + \sum_{t=2}^{\bar{T}}(Z_t-\rho Z_{t-1})(X_t-\rho X_{t-1})\} - \frac{f^{(1)}}{f}(W)(\bar{X}^f-E(\bar{X}^f)) \quad (3)$$
$$I = \sigma^{-2}\Sigma_1 + I_f\Sigma_2. \quad (4)$$

The proof is in the Appendix B.

We now construct an efficient estimator of β following the same ideas as in Park and Simar (1994). We need preliminary $\sqrt{N}$-consistent estimator $\tilde{\beta}$ and $\tilde{\rho}$ of β and ρ. The within estimator obtained by regressing $Y_{it}-\bar{Y}_i$ on $X_{it}-\bar{X}_i$ by OLS methods provides the consistent $\tilde{\beta}$ and the correlation of the within OLS residuals $Y_{it}-\bar{Y}_i-\tilde{\beta}(X_{it}-\bar{X}_i)$ with their lagged values provides a consistent $\tilde{\rho}$.

A consistent estimator of σ^2 can be derived as follows. Define $W_i(\rho;\beta);Z_{it}(\rho;\beta)$ and $\bar{X}^f_i(\rho)$ in the same way as defining $W(\rho;\beta);Z_t(\rho;\beta)$ and $\bar{X}^f(\rho)$ respectively. For example

$$W_i(\rho;\beta) = \sum_{t=1}^{\bar{T}} c_t(\rho)(Y_{it}-X_{it}'\beta).$$

Define

$$\hat{\sigma}^2(\rho;\bar{\mu}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} e_t(\rho)Z_{it}(\rho;\bar{\mu})^2;$$

where

$$e_t(\rho) = \begin{cases} (1+\rho)/(T-1) & ; \text{ for } t=1 \text{ and } T \\ (1-\rho^2)/(T-1) & ; \text{ for } t=2,\ldots,T-1 \end{cases}$$

Note that for $t=1,\ldots,T$

$$E\,Z_{it}(\rho;\bar{\mu})^2 = \frac{(T-1)(1-\rho)}{T(\rho)(1+\rho)}\sigma^2$$

$$e_t(\rho) = \frac{T(\rho)(1+\rho)}{(T-1)(1-\rho)}\,c_t(\rho);$$

and

$$\sum_{t=1}^{T} e_t = \frac{T(\rho)(1+\rho)}{(T-1)(1-\rho)}.$$

So we have $E(\hat{\sigma}^2(\rho;\bar{\mu})) = \sigma^2$. Now,

$$\hat{\sigma}^2 = \hat{\sigma}^2(\rho;\tilde{\mu}) \tag{5}$$

defines a consistent estimator of σ^2.

We need also to define an estimator of the matrix I: we define

$$\hat{I} = \sigma^{-2}\hat{S}_1 + \hat{I}_f\hat{S}_2;$$

where

$$\hat{S}_1 = \frac{1}{N}\sum_{i=1}^{N}\{(1-\rho^2)(X_{i1}-\hat{X}_i(\rho))(X_{i1}-\hat{X}_i(\rho))' + \sum_{t=2}^{T}[X_{it}-\hat{X}_i(\rho)-\rho(X_{i,t-1}-\hat{X}_i(\rho))][X_{it}-\hat{X}_i(\rho)-\rho(X_{i,t-1}-\hat{X}_i(\rho))]'\};$$

$$\hat{S}_2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{X}_i(\rho)-\hat{X}_c(\rho))(\hat{X}_i(\rho)-\hat{X}_c(\rho))'$$

$$\text{with } \hat{X}_c(\rho) = \frac{1}{N}\sum_{i=1}^{N}\hat{X}_i(\rho);$$

$$\hat{I}_f = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{f}'(1)}{\hat{f}}\right)^2(\hat{W}_i;\rho;\tilde{\mu});$$

In the last expression $\hat{W}_i = W_i(\rho;\tilde{\mu})$ and $\hat{f}$ is a kernel estimator of f, the pdf of W:

$$\hat{f}(w;\rho;\bar{\mu}) = \frac{1}{N}\sum_{i=1}^{N}K_s(w-W_i(\rho;\bar{\mu})) + c;$$

where $K_s(u) = (1/s)K(u/s)$, $K(u) = e^{iu}(1+e^{iu})^{-2}$ and the bandwidths and the constant c tends to zero at some appropriate rates described below. Finally, writing $\tilde{e}_{it} = Z_{it}(\tilde{\rho}, \tilde{\beta})$ we can define the estimator of β as follows:

$$\hat{\beta} = \tilde{\beta} + \frac{1}{N}\sum_{i=1}^{N} \hat{\sigma}_i^{-2}(1-\hat{\rho}^2)\tilde{e}_{i1}X_{i1}^{*} + \hat{\sigma}_i^{-2}\sum_{t=2}^{T}(\tilde{e}_{it} - \hat{\rho}\tilde{e}_{i,t-1})(X_{it}^{*} - \hat{\rho}X_{i,t-1}^{*}) - (\hat{f}_i(\hat{\rho}) - \hat{f}_c(\hat{\rho}))\frac{\hat{f}^{(1)}}{\hat{f}}(W_i, \hat{\rho}, \tilde{\beta})\}. \quad (6)$$

We state now our main results showing that $\hat{\beta}$ is a semiparametric efficient estimator of β:

Theorem 2.2 *Assume that $E(e^{t|e_1|}) < \infty$ for some $t > 0$ and that $\int |u|^2 h(u)\,du < \infty$. If $s \to 0$; $c \to 0$ and $Ncs^6 \to \infty$ as $N \to \infty$, then we have*

$$L_P(N^{1/2}(\hat{\beta} - \beta)) \to N(0, I^{-1}) \quad \text{as } N \to \infty.$$

The proof is in the Appendix B.

Remark 2.1 GLS within estimator
In the Theorem above, we used, $\tilde{\beta}$, the OLS within estimator of β as a first step consistent estimator. From this we derived also the consistent estimators $\hat{\rho}, \hat{\sigma}^2$. This is sufficient to define our semiparametric efficient $\hat{\beta}$ in (6). This first step within does not take into account the AR(1) structure of the error term. We could also use in the first step a feasible GLS within estimator of β which is more efficient than the OLS within. This might in small sample improve the quality of our semiparametric estimator. The GLS within can be defined as follows. Let

$$X_{i1}^{*} = X_{i1} - \tilde{X}_i(\rho)$$

and for $t = 2, \ldots, T$

$$X_{it}^{*} = X_{it} - \tilde{X}_i(\rho) - \rho(X_{i,t-1} - \tilde{X}_i(\rho)).$$

Likewise define Y_{it}^{*} for $t = 1, \ldots, T$. First, $\tilde{Y}_i(\rho) = \sum_{t=1}^{T} c_t(\rho)Y_{it}$ then

$$Y_{i1}^{*} = Y_{i1} - \tilde{Y}_i(\rho)$$

and for $t = 2, \ldots, T$

$$Y_{it}^{*} = Y_{it} - \tilde{Y}_i(\rho) - \rho(Y_{i,t-1} - \tilde{Y}_i(\rho)).$$

The GLS within estimator of β is then defined by

$$\tilde{\beta}_{GLS} = \Big(\sum_{i=1}^{N}\{(1-\rho^2)X_{i1}^{*}X_{i1}^{*\prime} + \sum_{t=2}^{T} X_{it}^{*}X_{it}^{*\prime}\}\Big)^{-1} \cdot \Big(\sum_{i=1}^{N}\{(1-\rho^2)X_{i1}^{*}Y_{i1}^{*} + \sum_{t=2}^{T} X_{it}^{*}Y_{it}^{*}\}\Big). \quad (7)$$

> From this estimator, as above we can derive an improved estimator of ρ and σ^2 by plugging $\tilde{\beta}_{GLS}$ in place of $\tilde{\beta}$ in the expressions defining ρ and then σ^2, as in (5). This provides ρ_{GLS} and σ^2_{GLS}. Finally all those GLS within estimators can be used in the expression (6) to provide $\hat{\alpha}_{GLS}$. In the simulations below, we will compare the small sample properties of $\hat{\alpha}$ and $\hat{\alpha}_{GLS}$.

Remark 2.2 *Another improved estimation of ρ and σ^2.*
We may re-estimate ρ with the efficient estimator of $\bar{\beta}$ at hand. For an improved estimator of ρ, we may consider the correlation of the residuals

$$Y_{it} - \bar{Y}_{i-1} - \hat{\alpha}(X_{it} - \bar{X}_{i-1})$$

with their lagged values, or the solution of the likelihood equation $\ell_\rho = 0$. This provides $\hat{\rho}$. Then, with $\hat{\rho}$ and $\hat{\alpha}$ we can also define by (5):

$$\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\rho}, \hat{\alpha}).$$

2.2 Model 2

Model 2 is in fact the general basic model specified by equation (1) where we allow dependency between α_i and X_i. We still assume independency between (α_i, X_i) and ε_i. Here we denote $h(\cdot,\cdot)$ the common $(1+dT)$-variate density of (α_i, X_i)'s. Then with all the notations introduced in Model 1, we can state the theorem:

Theorem 2.3 *The efficient score function and the information bound for estimating $\bar{\beta}$ in Model 2 are given by:*

$$\ell^* = \frac{1}{\sigma^2}\{(1-\rho^2)Z_1 X_1 + \sum_{t=2}^{T}(Z_t - \rho Z_{t-1})(X_t - \rho X_{t-1})\} \tag{8}$$

$$I = \sigma^{-2}\Sigma_1. \tag{9}$$

The proof is in the Appendix B.

Remark 2.3 *Obviously we have:*

$$\sigma^{-2}\Sigma_1^{-1} \cdot \sigma^{-2}\Sigma_1 + I_f\Sigma_2 \cdot E\,\ell^{-}\ell^{-0},$$

where the inequality between the matrices is in the sense of the positive semidefiniteness of their differences. The first inequality shows the price to pay for allowing dependency between α and X, and the next one for not knowing ρ, σ^2, h and g.

An semiparametric efficient estimator for Model 2 could be constructed as above in Model 1 but now, without the second part in $\hat{\alpha}$ (compare (3) with (8)), but a simpler efficient estimator is provided by the following corollary:

Corollary 2.1 *The GLS within estimator $\tilde{\beta}_{GLS}$ defined in (7) is efficient in Model 2.*

Proof. It is indeed straightforward to show that as $N \to \infty$, we have

$$\mathcal{L}_P(\sqrt{NT}(\tilde{\beta}_{GLS,N} - \beta)) \to N(0, \sigma^2 \Sigma_1^{-1}).$$

2.3 Individual effects and the level of the frontier.

Since $W_i = \alpha_i + \sum_{t=1}^{T} c_t^2 \varepsilon_{it}$, it seems natural to define as estimator of the effects:

$$\hat{\alpha}_i = W_i(\hat{\rho}, \hat{\beta}) \tag{10}$$

$\hat{\rho}$ is either ρ or its improvement discussed above. Let B be the upper boundary of the support of the marginal density of α_i's, and define its estimator by

$$\hat{B} = \max_{1 \le i \le N} W_i(\hat{\rho}, \hat{\beta}).$$

Assume X_{it}'s are i.i.d. d-dimensional vectors with $E\|X_{11}\|^2 < \infty$ and $E(X_{11} - E(X_{11}))(X_{11} - E(X_{11}))'$ being nonsingular. Suppose that both N and T go to infinity and that $\sqrt{NT}(\hat{\beta} - \beta) = O_p(1)$, $\sqrt{NT}(\hat{\rho} - \rho) = O_p(1)$. Then, we can show, as in Park and Simar (1994) or Park, Sickles and Simar (1998), that

$$\mathcal{L}_P(\sqrt{T}(\hat{\alpha}_i - \alpha)) \to N(0, \sigma^2/(1-\rho)^2)$$

and that

$$\hat{B} - B = O_p(T^{-1/2} \log N + N^{-1}).$$

3 Monte Carlo Simulations

We have 4 consistent estimators of β: the OLS within $\hat{\beta}$, the GLS within $\tilde{\beta}_{GLS}$ and our two efficient estimators $\hat{\alpha}$, $\tilde{\alpha}_{GLS}$. The finite sample peformances are compared through the following Monte-Carlo (MC) scenarios:

We simulate samples of size $n = 20; 100; 1000$ with $t = 12; 60$ in a model with $d = 2$ regressors. In each MC sample, the regressors are generated according a bivariate VAR model:

$$X_{it} = R X_{i;t-1} + \eta_{it}, \quad \text{where } \eta_{it} \sim IN_2(0, \sigma_X^2 I_2),$$

where $\sigma_X = 1$ and $R = \begin{pmatrix} 4 & 0.5 \\ 0.5 & 4 \end{pmatrix}$. Then the obtained values of X_{it} were shifted arround three different means to obtain almost 3 balanced groups of firms from smaller to larger. We fixed $\mu_1 = (5 \quad 5)', \mu_2 = (7.5 \quad 7.5)', \mu_1 = (10 \quad 10)'$. The idea is to generate a raisonnable cloud of points for X. Other scenarios have been tried: they influence the quality of the estimators jointly but they do not change the conclusions on the comparison issue raised here.

The autoregressive $AR(1)$ part of the model was generated with $\rho = 0.7; 0.1$ and $\sigma = 0.5$. For small values of ρ we could expect that finite sample performances of our efficient estimator could be questionable. Changing the value of σ would of course affect jointly the quality of all the estimators but does not affect the comparisons done below.

Finally, the inefficiency part (the individual effects) are generated independently of the regressors as $B \sim Expo(\alpha)$ where we choosed for the exponential distribution a mean $\alpha = 1$ and for the upper boundary a value of $B = 1$. Since they are often measured in logarithms (like in Cobb Douglas production functions), this involves an average inefficiency score of 0.5. Here again, other scenarios for generating the α_i could be chosen but this does not affect the conclusions below. The values of β was set equal to $(1 \quad 0.5)'$.

Due to computing time limitations, most of the results are done with $M = 500$ MC replications but when $n = 1000$ only $M = 100$ replications were performed. Some scenarios (with smaller n) were done with $M = 1000$ confirming the reported results.

Since the VAR process generating the regressors X_i is symmetric in both components, the MSE for the estimators of the two coefficients are of the same order of magnitude. So, to summarize the results, the Tables display the sum of the two MC Mean squarred errors:

$$MSE = \sum_{j=1}^{2} \frac{1}{M} \sum_{m=1}^{M} (\hat{\beta}_j^m - \beta_j)^2;$$

where $\hat{\beta}$ is one of the four proposed estimators.

For the bandwiths we selected an optimal fixed values s^* by running the whole Monte-Carlo experiment for a selected grid of 20 equally spaced values for s between 0.1 to 2. We report in the Tables the results corresponding to the optimal bandwidth s^*, where optimal means which minimized MSE. In all the tried scenarios, the results were not very sensitive to the choice of s in the above grid.

In a following section below we will propose a data driven method to derive an optimal bandwidth s^* with a real data set: the idea will be to replace the Monte-Carlo experiment by a Bootstrap algorithm.

In the situation of Table 1, $\rho = 0.7$: with such a value for the autoregressive parameter, we see a clear improvement in the use of the efficient estimator, in particular, compared with

n	t	$\tilde{\beta}$	$\tilde{\beta}_{GLS}$	$\hat{\beta}$	$\hat{\beta}_{GLS}$	s^2
20	12	42.64	20.19	17.92	17.91	0.4
100	12	8.523	4.076	3.591	3.621	0.3
1000	12	0.7527	0.3408	0.3031	0.3086	0.8

Table 1: Monte-Carlo MSE of the estimators of β with $M = 500$ replications. The figures for the MSE are multiplied by 10^4. Here $\rho = 0.7$, $\phi = 0.5$ and $\sigma^2_\alpha = 1$. For $n = 1000$ only $M = 100$ replications were performed.

the OLS within. Note also how the GLS within dominates the OLS. By looking to the two versions of our efficient estimator, it is not clear that we gain by chosing the GLS within as a first step for defining the estimator: this is confirmed in most of the scenarios below. When increasing the value of t we better estimate ρ and the improvement is still better as confirmed in Table 2 with $t = 60$.

n	t	$\tilde{\beta}$	$\tilde{\beta}_{GLS}$	$\hat{\beta}$	$\hat{\beta}_{GLS}$	s^2
20	60	11.41	3.544	3.492	3.494	0.9
100	60	2.122	0.7102	0.6933	0.6931	0.3
1000	60	0.2104	0.0597	0.0584	0.0585	0.4

Table 2: Monte-Carlo MSE of the estimators of β with $M = 500$ replications. The figures for the MSE are multiplied by 10^4. Here $\rho = 0.7$, $\phi = 0.5$ and $\sigma^2_\alpha = 1$. For $n = 1000$ only $M = 100$ replications were performed.

When the autocorrelation ρ is small we might expect poor performances, in finite samples, of our efficient estimators, since the correction factor in (6) introduces additional noise. This is invetigated in Table 3 where $\rho = 0.1$. The Table shows that our efficient estimator behaves pretty well and better than the GLS within. The latter is not significantly different from the OLS. A better estimation of ρ, by increasing t does not change substantially these comments, as shown in Table 4.

n	t	$\tilde{\beta}$	$\tilde{\beta}_{GLS}$	$\hat{\beta}$	$\hat{\beta}_{GLS}$	s^2
20	12	23.15	23.31	20.12	20.15	0.3
100	12	4.464	4.496	3.623	3.622	0.2
1000	12	0.4026	0.4016	0.3398	0.3398	0.3

Table 3: Monte-Carlo MSE of the estimators of β with $M = 500$ replications. The figures for the MSE are multiplied by 10^4. Here $\rho = 0.1$, $\phi = 0.5$ and $\sigma^2_\alpha = 1$. For $n = 1000$ only $M = 100$ replications were performed.

n	t	$\bar{\alpha}$	$\bar{\alpha}_{GLS}$	$\hat{\alpha}$	$\hat{\alpha}_{GLS}$	s^2
20	60	3.6622	3.5535	3.4998	3.4998	0.9
100	60	0.7428	0.7222	0.7028	0.7028	0.3
1000	60	0.0000	0.0000	0.0000	0.0000	0.0

Table 4: Monte-Carlo MSE of the estimators of $\bar{\alpha}$ with $M = 500$ replications. The figures for the MSE are multiplied by 10^4. Here $\rho = 0.1$, $\sigma = 0.5$ and $\sigma_\alpha = 1$. For $n = 1000$ only $M = 100$ replications were performed.

The good performances of the efficient estimator with weak autocorrelation are confirmed by Table 5 where the data were generated with no autocorrelation ($\rho = 0$). We only display the results with small sample sizes to save space. Note that here, the GLS within introduces additional noise probably due to the estimation of ρ.

n	t	$\bar{\alpha}$	$\bar{\alpha}_{GLS}$	$\hat{\alpha}$	$\hat{\alpha}_{GLS}$	s^2
20	12	20.336	20.80	18.33	18.31	0.4
100	12	3.879	3.902	3.323	3.323	0.2

Table 5: Monte-Carlo MSE of the estimators of $\bar{\alpha}$ with $M = 500$ replications. The figures for the MSE are multiplied by 10^4. Here $\rho = 0$, $\sigma = 0.5$ and $\sigma_\alpha = 1$.

As a global conclusion, it appears that our efficient estimator behaves pretty well across the different MC scenarios even if ρ is small. When autocorrelation is present and more important, it definitely increases the precision of the estimators of $\bar{\alpha}$ for the different sample sizes analysed here.

To compute the efficient estimator $\hat{\alpha}$ we need to choose a bandwidth s. We propose here the following bootstrap procedure. For a given s we compute $\hat{\alpha}(s)$ and the resulting improved estimators $\hat{\beta}_i(s)$, $\hat{\rho}(s)$ and $\hat{\sigma}^2_\varepsilon(s)$. Then by using a semiparametric bootstrap, we generate B pseudo-samples $(X_{it}, Y^*_{it})^b$, $b = 1, \ldots, B$ by resampling on the residuals in model (1):

$$\begin{aligned} Y^*_{it} &= X'_{it}\hat{\alpha}(s) + \hat{\beta}_i(s) + \varepsilon^*_{it}, \quad i = 1, \ldots, N; \; t = 1, \ldots, T \\ \varepsilon^*_{it} &= \hat{\rho}(s)\varepsilon^*_{i,t-1} + u^*_{it}, \end{aligned} \tag{11}$$

where $u^*_{it} \sim N(0, \hat{\sigma}^2_\varepsilon(s))$.

Each pseudo-sample provides $\hat{\alpha}^{*b}(s)$. We compute the criterion value:

$$C(s) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\alpha}^{*b}(s) - \hat{\alpha}(s)\right)'\left(\hat{\alpha}^{*b}(s) - \hat{\alpha}(s)\right).$$

The optimal bandwith choice is then given by $s^* = \arg\min_s C(s)$. In the example we chose $B = 500$ and we carried out a grid search for s^* on 10 equally spaced values from 0.1 to 2.

1.

The optimal value is $s^* = 0.9$ and Table 6 summarizes the obtained results.

variable	OLS within $\tilde{\beta}$		GLS within $\tilde{\beta}_{GLS}$		Efficient $\hat{A}$	
X_1	0.96603	(0.01770)	0.98993	(0.01661)	0.99445	(0.01688)
X_2	0.48850	(0.01164)	0.49114	(0.01162)	0.49059	(0.01161)

Table 6: Estimation of β in our example where $n = 100$, $t = 12$. Here $\hat{\sigma}_\alpha^2 = 0.55445$ and $\hat{\sigma}_\nu^2 = 0.55445$. In parenthesis are the estimated standard deviations, for the OLS within they are the standard ones computed under the hypothesis that $\lambda = 0$.

Once the efficient estimators of β are obtained, estimators of λ, of the effects α_i and of the frontier level B may be derived. This will be illustrated with real data in the next example.

4 Empirical Illustration

4.1 Regulatory Issues in the U.S. Banking Industry

A comprehensive account of deregulation measures affecting capital requirements, product lines and geographic expansion is presented in Berger, Kashyap and Scalise (1995). The early banking legislature was aimed at protecting consumers from exploitation by regional and national money trusts. However the preservation of local monopolies and oligopolies in community banking was an unfortunate outcome. Since the early 1980s federal and state regulatory agencies have resorted to less stringent interpretation of banking regulations and adopted less restrictive legislature. The passing of the Reigle-Neal Act in the early 1990s enabled nationwide banking, while the relaxing of unit bank, branch bank and state bank type legislature have resulted in numerous mergers and failures which have significantly altered the US banking environment (Adams, Berger and Sickles 1997). The introduction of interest bearing consumer checking accounts and the phasing out of Regulation Q interest rate ceilings on savings and small denomination time deposits in the early 80s were among the initial wave of deregulation policies. Money market deposit accounts (structured similar to mutual funds) led not only to a new product line but also to competition from non-bank institutions.

Technical and …nancial innovations have contributed in transforming (and continues to transform) the US banking environment by eliminating natural barriers to expansion and by undermining statutory restraints.[2] Improvements in communication technology, better computers, extensive use of ATMs (automated teller machines) and new services (on-line banking) have played a key role in rede…ning this industry as we know it today. In addition, innovations in applied …nance, information processing and communications have fostered greater expansion of banking powers through diversi…cation and product synergies. These changes in the US banking environment have aroused considerable interest in the modeling and measuring of competition. A survey by Humphrey and Pulley (1997) presents a description of the deregulation measures implemented and analyzes the industry's reactions and adjustments.

The di¢culties associated in de…ning a bank's inputs, outputs and measuring prices have resulted in a substantial portion of the bank competition research being centered on the structure-performance relationship and related hypothesis. According to the structure-performance hypothesis the level of competition among …rms is in‡uenced by the degree of concentration within the industry.[3] The structural consolidation of the US banking sector in the recent past has fueled traditional fears of non-competitiveness. This hypothesis views that greater market concentration is conducive for more e¤ective collusion. The papers by Berger and Hannan (1989, 1998) and Berger (1995) provide an extensive overview of the prior research that used the structure-performance paradigm. The previous research referred to in the above papers has relied on reduced form models to measure market conduct. Bank performance measures such as return on equity (ROE), return on assets (ROA), interest rates charged on loans and interest rates paid on deposits were regressed on concentration measures, and these parameter estimates were used as indicators of the degree of competition among banks. Berger and Hannan (1989) used the e¢cient-structure hypothesis to measure competition by considering the e¤ect of market concentration on pro…tability. The vast literature on the above hypotheses generally support the structure-performance paradigm and partially supports the e¢cient-structure paradigm, with neither one completely consistent with the empirically observed relationships among pro…ts, prices, market structure and e¢ciency. Structural models designed to deal with these issues recently have been analyzed by Adams, Roeller, and Sickles (1999), Adams, Bauer, and Sickles (1999), and Adams, Jayasuriya, and Sickles (1999). We do pursue such structural modeling approaches in this paper.

[2] See Berger, Kashyap and Scalise (1995).

[3] HHI is widely used as a measure of market concentration. See Hannan (1997) for a discussion on the usefulness of the HHI index in the pricing of deposit and loan rates in the US banking industry.

4.2 Data[4]

A panel data set from the first quarter 1980 till the fourth quarter 1996 consisting of 10,000 US banks was used in the estimation of these models.

The production and cost data for the US banking industry was obtained on-line from the Federal Reserve Bank of Chicago. The Report of Condition and Income (Call Report), and the FDIC Summary of Deposits are the primary sources for the US banking data.

This panel data set is a comprehensive source of information on operating costs, inputs (including labor, capital and purchase funds), outputs (loans and deposit services), assets, and the regulatory environment of any institution in the US banking industry. Data on over one hundred variables was collected from the Call Reports and the FDIC Summary of Deposits.[5]

Labor (LAB) is measured using the number of full time-equivalent employees on the payroll at the end of each quarter. The total value of premises, fixed assets, and capitalized leases are used as a proxy for capital (CAP). Purchase funds (PURF) are measured using the sum of deposits greater than US$ 100,000, foreign debt, federal funds purchased, and liabilities on borrowed money.

The measurement of loan and deposit services is a more complex issue, and two approaches are currently utilized in the US banking literature: intermediation approach and production approach. The intermediation approach uses the dollar amounts of deposits and outstanding loans as a proxy for deposit and loan services provided by a bank, while the production approach uses the number of outstanding loans and deposits as a measure of banking services produced. The former approach is followed in the data collection and in the modeling method.

---

[4] The data sets used in this paper were compiled by Ruwan Jayasuriya while employed as a Summer Intern at the Antitrust Division, U.S. Department of Justice, Washington DC. The insightful comments by numerous economists working in the Antitrust Division, and the financial support provided are gratefully acknowledged.

[5] A detailed description of these variables can be found at the Federal Reserve Bank of Chicago URL (http://www.frbchi.org).

Production and Cost Data Description		
Variable	Mean	Std. Dev
Real Estate Loans (RELN)	50 million	23 million
Installment Loans (INLN)	17 million	11 million
Comm & Indus Loans (CILN)	18 million	17 million
Deposits	20,405	12,602
Price of RELN	3.92	1.08
Price INLN	4.34	1.05
Price of CILN	3.29	1.15
Price of Deposits	3.82	1.24
Labor	2,047	1,300
Capital	29 million	19 million
Purchase Funds	14 million	9 million
Wage	20,227	6,203
Price of Capital	13.38	3.81
Price of Purchase Funds	3.29	1.24

The following loan and deposit types are used in this study: real estate loans (RELN), commercial and industrial loans (CILN), installment loans (INLN), and retail time and savings deposits (Deposits).[6] CILN accounts for loans given to businesses, while INLN accounts for loans given to individuals to meet medical expenses, vacation expenses, purchase furniture, automobiles, household appliances, and other miscellaneous expenses. RELN accounts for loans secured by real estate.

The price (interest rate) for each of the loan types is obtained by dividing the interest rate and fee income earned, by the outstanding loan amount. A composite wage rate is obtained by dividing the total labor expenses by the total number of workers. Price indices for capital and purchase funds are calculated by dividing the expenses incurred for each input by the value of total deposits (as presented below).

$$w_j = \frac{\text{expenditure on input } j}{\text{total deposits}}$$

Outputs, inputs and price definitions used in this paper are consistent with those used in previous studies (Berger (1991), and Berger, Hancock and Humphrey (1993)). Bank size (total assets) is highly correlated with the size of a given output, and thus dollar values are used in place of the number of loans or deposits

[6] Retail time and savings deposits greater than US$ 100,000 are included in purchase funds.

The definitions of quantities and prices are less than ideal, but are necessitated due to the absence of explicit price indices. The Call Report and FDIC data are reported in nominal terms, and are converted into real terms using a state level consumer price index (1982-84 = $100).

4.3 Results

5 Conclusion

Appendix

A Efficient Estimation in Semiparametric Models

The notion of efficient estimation in semiparametric models is well established in Begun, Hall, Huang and Wellner (1983), Bickel, Klaassen, Ritov and Wellner (1993), and Pagan and Ullah (1994). An excellent survey may be found in Newey (1990). Below we briefly outline the basic ideas in our context.

Write (X, Y) for "generic" observations. Let $\mathbb{P}$ be the set of all possible joint distributions of (X, Y). Let θ be the parameter vector. One calls $\mathbb{P}_0$ a regular parametric submodel of $\mathbb{P}$ if $\mathbb{P}_0 (\subset \mathbb{P})$ can be represented as $\{P_{(\theta,\eta)} : \theta \in \mathbb{R}^d, \eta \in S \text{ open} \subset \mathbb{R}^k\}$ and at every (θ_0, η_0) the mapping $(\theta, \eta) \to P_{(\theta,\eta)}$ is continuously Hellinger differentiable (see Ibragimov and Has'minskii, 1981, Section 1.7).

For instance, in our basic model (1), a probability distribution can be characterized by (θ, h, σ^2). Let $(\theta_0, h_0, \sigma_0^2)$ denote the true value of the true underlying P and consider a class of functions $h_\eta(\cdot)$ indexed by $\eta_1 \in \mathbb{R}^1$ where h_η is identical to the true h_0 when $\eta_1 = 0$ (for example, one may take $h_\eta(\cdot) = h_0(\cdot + \eta_1)$). Write η_2 for σ^2, with $\eta_{20} = \sigma_0^2$. Let $P_{(\theta,\eta_1,\eta_2)}$ denote a distribution characterized by (θ, η_1, η_2). This means that the true P can be written as $P_{(\theta_0,0,\eta_{20})}$. Then, the class of probability distribution $\{P_{(\theta,\eta_1,\eta_2)} : \theta \in \mathbb{R}^d, \eta_1 \in \mathbb{R}^1, \eta_2 > 0\}$ is a submodel of $\mathbb{P}$ passing through the true P and is regular if h_{η_1}, as functional of η_1, is "smooth" in a certain sense (Ibragimov and Has'minskii, 1981).

Suppose $P (= P_{(\theta_0,\eta_0)})$ belongs to a regular parametric submodel $\mathbb{P}_0$ of $\mathbb{P}$. Then the notion of information bound and efficient estimation of θ are well defined.

Let $L(X, Y; \theta, \eta)$ denote the log likelihood of an observation from $P_{(\theta,\eta)}$ and let $\ell_\theta(X, Y) = \partial L / \partial \theta |_{(\theta_0,\eta_0)}$ and $\ell_{\eta_j}(X, Y) = \partial L / \partial \eta_j |_{(\theta_0,\eta_0)}$ where $\eta = (\eta_1, \ldots, \eta_k)$. Then,

$$I(P; \theta; \mathbb{P}_0) = E[\ell_\theta - \sum_{j=1}^{k} c_j^* \ell_{\eta_j}][\ell_\theta - \sum_{j=1}^{k} c_j^* \ell_{\eta_j}]'$$

where c_j^* is a d-dimensional vector uniquely determined by the orthogonality condition:

$$E[\ell_\theta - \sum_{j=1}^{k} c_j^* \ell_{\eta_j}]\ell_{\eta_j} = 0 \quad j = 1, \ldots, k$$

In fact, the information $I(P; \theta; \mathbb{P}_0)$ given above is nothing else than the inverse of $d \times d$ top-left partition of $[E(\ell\ell')]^{-1}$ where $\ell = (\ell_\theta, \ell_{\eta_1}, \ldots, \ell_{\eta_k})'$.

Moreover, it can be also written as

$$I(P; \theta; \mathbb{P}_0) = E\ell^* \ell^{*\prime}$$

where

$$\dot{\ell}^{*} = \dot{\ell} - \Pi(\dot{\ell} \mid [\dot{\ell}_\eta]),$$

$[\dot{\ell}_\eta]$ denotes the linear span generated by $\{\dot{\ell}_{\eta j}\}_{j=1}^{k}$, and $\Pi(\dot{\ell} \mid S)$ denotes the vector of projections of each component of $\dot{\ell}$ onto the space S in $L_2(P)$. In other words, we project the scores with respect to the slope parameters onto the nuisance parameter tangent space and then purge the scores of these projections to get the efficient score, which is then orthogonal to the nuisance parameters. An estimator of ψ is called efficient if it is asymptotically normal with mean zero and variance $N^{-1} I^{-1}(P; \psi; \mathbb{P}_0)$.

The above discussion applies when P ranges over $\mathbb{P}_0$. Clearly, if we only assume that $P \in \mathbb{P}$ we can estimate no better than if we assumed that $P \in \mathbb{P}_0$. Accordingly, let $I(P; \psi; \mathbb{P}) = \inf\{I(P; \psi; \mathbb{P}_0) : \mathbb{P}_0$ is a regular parametric submodel of $\mathbb{P}$; $P \in \mathbb{P}_0\}$ be the information bound for estimating ψ under $\mathbb{P}$. An estimator $\hat{\psi}_N$ is now called efficient in $\mathbb{P}$ if

$$L_P(\sqrt{N}(\hat{\psi}_N - \psi)) \Rightarrow N(0, I^{-1}(P; \psi; \mathbb{P})).$$

A method of finding $I(P; \psi; \mathbb{P})$ is well explained in Bickel, Klaassen, Ritov and Wellner (1993). Let C denote the class of all regular parametric submodels containing P, and let $[\dot{\ell}(\mathbb{P}_0)]$ denote the linear span generated by $\dot{\ell}$ for a submodel $\mathbb{P}_0$. Then $I(P; \psi; \mathbb{P})$ can be obtained by

$$I(P; \psi; \mathbb{P}) = E\,\dot{\ell}^{*}\dot{\ell}^{*\prime}(X, Y)$$

where

$$\dot{\ell}^{*} = \dot{\ell} - \Pi(\dot{\ell} \mid V)$$

and V is the closed linear span (called tangent space) of the union of $[\dot{\ell}(\mathbb{P}_0)]$ when $\mathbb{P}_0$ ranges over C. The random variable $\dot{\ell}^{*}$ is called the efficient score function.

B Lemmas and Proofs

Proof of Theorem 2.1.

References

[1] Adams, R.M., Berger, A.N., and Sickles, R.C. (1998), "Semiparametric Approaches to Stochastic Panel Frontiers with Applications in Applications in the Banking Industry," Journal of Business and Economic Statistics, forthcoming.

[2] Adams, R.M., Berger, A.N., and Sickles, R.C. (1997b), "Computation and Inference in Semiparametric Efficient Estimation," Advances in Computational Economics, edited by Hans Amman, Berc Rustem and Andrew Whinston, Boston: Kluwer Academic, 57-70..

[3] Arrelano, M. and Bover, D. (1995) "Another look at the instrumental variable estimation of error-components models," Journal of Econometrics 68,29–51.

[4] Begun, J.M., Hall, W.J., Huang, W.M., and Wellner, J.A. (1983), Information and asymptotic efficiency in parametric-nonparametric models. Annals of Statistics, 11, 432-452.

[5] Berger, Allen N. (1993), "'Distribution-Free' Estimates of Efficiency in U.S. Banking Industry and Tests of the Standard Distributional Assumptions," Journal of Productivity Analysis 4, 261–92.

[6] Berger, Allen N., Kashyap, Anil K., Scalise Joseph M. (1995), "The Transformation of the US Banking Industry: What a Long Strange Tips It's Been" Brookings Papers on Economic Activity 2, 55–218.

[7] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993), Efficient and Adaptive Estimation in Non- and Semi-parametric Models. Baltimore: Johns Hopkins University Press.

[8] Chamberlain, G. (1987) "Asymptotic Efficiency in Estimation with Conditional Moments Restrictions," Journal of Econometrics 34,305–34.

[9] Chamberlain, G. (1992) "Efficiency Bounds for Semiparametric Regressions," Econometrica 60, 567–96.

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

[11] Diewert, W.E. (1982), "Duality Approaches to Microeconomic Theory," in Arrow Kenneth J. and intrilligator, Michael D. (eds.), Handbook of Mathematical Economics, Volume II. Amsterdam: North Holland.

[12] Hausman, J.A. and Taylor, E.E. (1981), "Panel Data and Unobservable Individual Effects," Econometrica, 49, 1377–98.

[13] Horowitz, J.L. and Markatou, M. (1996), "Semiparametric Estimation of Regression Models for Panel Data," Review of Economic Studies 63, 145-168.

[14] Horrace, W. (1997) "A Semiparametric Error Component Density Estimation Technique for Stochastic Frontier Models," mimeo, University of Arizona.

[15] Ibragimov, I.A. and Has'minskii, R.Z. (1981), Statistical Estimation: Asymptotic Theory. Springer, New-York.

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

[17] Newey, W.K. (1990), Semiparametric efficiency bounds. Journal of Applied Econometrics, 5, 99–135.

[18] Pagan, P., and Ullah, A. (1994) Non-parametric Econometrics, mimeo.

[19] Park, B.U. and Simar, L. (1995), "Efficient Semiparametric Estimation in Stochastic Frontier Models," Journal of the American Statistical Association 89, 929–36.

[20] Park, B.U., Sickles, R.C. and Simar, L (1998), "Stochastic Frontiers: A Semiparametric Approach," Journal of Econometrics, forthcoming.

[21] Robinson, P. (1988), "Root-N-consistent Semiparametric Regression," Econometrica 56, 931–54.

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

[23] Scott, David W. (1992), Multivariate Density Estimation: Theory, Practice, and Visualization. New York: John Wiley & Sons.

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