Semiparametric Efficient Estimation of AR(1) Panel Data Models*

Byeong U. Park†
Department of Statistics
Seoul National University

Robin C. Sickles
Department of Economics
Rice University

Léopold Simar‡
Institut de Statistique
Université Catholique de Louvain

August 30, 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). 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 (1999). Theoretical derivations are supplemented by Monte Carlo simulations. We also provide an empirical illustration by estimating relative efficiencies from a stochastic distance function for the U. S. banking industry over the 1980's and 1990's. In markets where regulatory constraints have been lessened or done away with, the deregulatory dynamic market shocks may not be adjusted to immediately and may induce a serial correlation pattern in firm's use of best-practice banking technologies. Our semiparametric estimators have an important role in providing robust point estimates and inferences of the productivity and efficiency gains due to such economic reforms.

---

*The authors would like to thank Wonho Song for his valuable research assistance.

†Research support by the Grant No. 1999-1-104-001-5 from the interdisciplinary research program of the KOSEF and by Brain Korea 21 Project is acknowledged.

‡Research support from "Projet d'Actions de Recherche Concertées" (No. 98/03–217) from the Belgian Government is acknowledged.

1 Introduction

This study focuses on the semiparametric efficient estimation of random effect panel models containing AR(1) disturbances as well as generalizations to models in which the effects and regressors are correlated (Hausman and Taylor, 1981). 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. These estimators extend the previous work of Park and Simar (1994), Park, Sickles, and Simar (1998), and Adams, Berger, and Sickles (1999).

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 (1994) 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 explored 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. 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 (1999) developed semiparametric efficient estimators using higher dimensional 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 estimators derived by Park and Simar (1994) and by Park, Simar, and Sickles (1998) by allowing for the transitory error to be autocorrelated.

One motivation for such a model which we illustrate in an empirical example is the need to estimate an serially correlated 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, the 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, semiparametric efficient estimation becomes an important tool in providing robust point estimation of inferences of financial deregulation's impact on productivity and efficiency.

Section 2 contains our main results. There we outline our general panel model and 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 as well as the results from our application in analyzing banking productivity. Section 5 concludes. All 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:

$$\begin{aligned} Y_{it} &= X_{it}'\beta + \alpha_i + \varepsilon_{it} \quad ; \quad i = 1, \ldots, N; t = 1, \ldots, T \\ \varepsilon_{it} &= \rho\varepsilon_{i,t-1} + u_{it} \quad ; \quad |\rho| < 1 \end{aligned} \tag{1}$$

where $X_{it} \in \mathbb{R}^d$, $\beta \in \mathbb{R}^d$ and u_{it} are iid random variables from a $\mathcal{N}(0, \sigma^2)$. Denoting $X_i = (X_{i1}', \ldots, X_{iT}')'$, (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).

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 and terminologies 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}), $i = 1, \ldots, N$. Define

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

where $T(\rho) = (1-\rho^2) + (T-1)(1-\rho)^2$ and $v^2 = v^2(\rho) = \sigma^2 / T(\rho)$. The pdf of (X, Y) can be written as

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

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

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

where the second term follows a $\mathcal{N}(0, v^2)$ distribution. Thus the pdf of W is given by

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

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

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

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 define

$$\begin{aligned}\Sigma_1 &= E\{(1-\rho^2)(X_1 - \widetilde{X})(X_1 - \widetilde{X})' \\ &\quad + \sum_{t=2}^{T}(X_t - \widetilde{X} - \rho(X_{t-1} - \widetilde{X}))(X_t - \widetilde{X} - \rho(X_{t-1} - \widetilde{X}))'\}, \\ \Sigma_2 &= E(\widetilde{X} - E\widetilde{X})(\widetilde{X} - E\widetilde{X})'.\end{aligned}$$

We can now state the following theorem:

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

$$\ell^* = \frac{1}{\sigma^2}\{(1-\rho^2)Z_1X_1 + \sum_{t=2}^{T}(Z_t - \rho Z_{t-1})(X_t - \rho X_{t-1})\} - \frac{f^{(1)}}{f}(W)(\widetilde{X} - E(\widetilde{X})) \quad (3)$$

$$I = \sigma^{-2}\Sigma_1 + I_f\Sigma_2. \quad (4)$$

The proof is in the Appendix B.

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

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

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

Define

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

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 $t = 1, \dots, T$

$$E\ Z_{it}(\rho,\beta)^2 = \frac{(T-1)(1-\rho)}{T(\rho)(1+\rho)}\,\sigma^2$$
$$e_t(\rho) = \frac{T(\rho)(1+\rho)}{(T-1)(1-\rho)}\,c_t(\rho),$$
$$\sum_{t=1}^{T} e_t(\rho) = \frac{T(\rho)(1+\rho)}{(T-1)(1-\rho)}.$$

So we have $E(\hat\sigma^2(\rho,\beta)) = \sigma^2$. A consistent estimator of σ^2 is given by

$$\tilde\sigma^2 = \hat\sigma^2(\tilde\rho,\tilde\beta). \tag{5}$$

We need also to define an estimator of the matrix I. To do this let $\widetilde{X}_{\cdot}(\tilde\rho) = \frac{1}{N}\sum_{i=1}^{N}\widetilde{X}_i(\tilde\rho)$, and write

$$\begin{aligned}
\widehat\Sigma_1 &= \frac{1}{N}\sum_{i=1}^{N}\{(1-\tilde\rho^2)(X_{i1}-\widetilde{X}_i(\tilde\rho))(X_{i1}-\widetilde{X}_i(\tilde\rho))' \\
&\quad+\sum_{t=2}^{T}[X_{it}-\widetilde{X}_i(\tilde\rho)-\tilde\rho(X_{i,t-1}-\widetilde{X}_i(\tilde\rho))][X_{it}-\widetilde{X}_i(\tilde\rho)-\tilde\rho(X_{i,t-1}-\widetilde{X}_i(\tilde\rho))]'\}, \\
\widehat\Sigma_2 &= \frac{1}{N}\sum_{i=1}^{N}(\widetilde{X}_i(\tilde\rho)-\widetilde{X}_{\cdot}(\tilde\rho))(\widetilde{X}_i(\tilde\rho)-\widetilde{X}_{\cdot}(\tilde\rho))' \\
\widehat{I}_f &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat f^{(1)}}{\hat f}\right)^2(\widetilde{W}_i;\tilde\rho,\tilde\beta).
\end{aligned}$$

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

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

where $K_s(u) = (1/s)K(u/s)$, $K(u) = e^{-u}(1+e^{-u})^{-2}$ and the bandwidth s and the constant c tends to zero at some appropriate rates described below. We define $\widehat{I} = \tilde\sigma^{-2}\widehat\Sigma_1 + \widehat{I}_f\widehat\Sigma_2$.

Finally, writing $\widetilde{Z}_{it} = Z_{it}(\tilde\rho,\tilde\beta)$ we can define the estimator of β by

$$\begin{aligned}
\hat\beta &= \tilde\beta + \frac{1}{N}\widehat{I}^{-1}\sum_{i=1}^{N}\{\tilde\sigma^{-2}(1-\tilde\rho^2)\widetilde{Z}_{i1}X_{i1} \\
&\quad+\tilde\sigma^{-2}\sum_{t=2}^{T}(\widetilde{Z}_{it}-\tilde\rho\widetilde{Z}_{i,t-1})(X_{it}-\tilde\rho X_{i,t-1}) - (\widetilde{X}_i(\tilde\rho)-\widetilde{X}_{\cdot}(\tilde\rho))\frac{\hat f^{(1)}}{\hat f}(\widetilde{W}_i;\tilde\rho,\tilde\beta)\}.
\end{aligned} \tag{6}$$

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

Theorem 2.2 *Assume that $E(e^{t|\widetilde{X}_1|}) < \infty$ for some $t > 0$ and that $\int |u|^2 h(u)\,du < \infty$. If $s \to 0, c \to 0$ and $Nc^2s^6 \to \infty$ as $N \to \infty$, then we have*

$$\mathcal{L}_P(N^{1/2}(\hat\beta-\beta)) \to \mathcal{N}(0, I^{-1}) \qquad \text{as } N \to \infty.$$

The proof is in Appendix B.

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 $\tilde{\rho}, \tilde{\sigma}^2$. 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 terms. 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

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

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

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

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

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

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

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

The **GLS within** estimator of β is then defined by

$$\begin{aligned} \tilde{\beta}_{GLS} &= (\sum_{i=1}^{N} \{(1-\tilde{\rho}^2) X_{i1}^* X_{i1}^{*\prime} + \sum_{t=2}^{T} X_{it}^* X_{it}^{*\prime}\})^{-1} \\ &\quad \times (\sum_{i=1}^{N} \{(1-\tilde{\rho}^2) X_{i1}^* Y_{i1}^* + \sum_{t=2}^{T} X_{it}^* Y_{it}^*\}). \end{aligned} \tag{7}$$

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 $\tilde{\rho}$ and then $\tilde{\sigma}^2$, as in (5). This provides $\tilde{\rho}_{GLS}$ and $\tilde{\sigma}^2_{GLS}$. Finally all those GLS within estimators can be used in the expression (6) to provide $\hat{\beta}_{GLS}$. In the simulations below, we will compare the small sample properties of $\hat{\beta}$ and $\hat{\beta}_{GLS}$.

Remark 2.1 Another improved estimation of ρ and σ^2.
We may re-estimate ρ with the efficient estimator of β at hand. For an improved estimator of ρ, we may consider the correlation of the residuals

$$Y_{it} - \bar{Y}_i - \hat{\beta}(X_{it} - \bar{X}_i)$$

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

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

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 following theorem.

Theorem 2.3 *The efficient score function and the information bound for estimating β in Model 2 are given by:*

$$\ell^* = \frac{1}{\sigma^2}\{(1-\rho^2)Z_1X_1 + \sum_{t=2}^{T}(Z_t - \rho Z_{t-1})(X_t - \rho X_{t-1})\} \tag{8}$$

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

The proof is in Appendix B.

Remark 2.2 *Obviously we have:*

$$\sigma^{-2}\Sigma_1 \leq \sigma^{-2}\Sigma_1 + I_f\Sigma_2 \leq E\ell_\beta\ell'_\beta,$$

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.

A semiparametric efficient estimator for *Model 2* could be constructed as above in *Model 1* but now, without the second part in ℓ^* (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

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

2.3 Individual effects and the level of the frontier.

Since $W_i = \alpha_i + \sum_{t=1}^{T} c_t\epsilon_{it}$, it seems natural to define estimators of the effects by

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

where $\hat{\rho}$ is either $\tilde{\rho}$ 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

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

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 $\sqrt{NT}(\tilde{\beta} - \beta) = O_p(1)$, $\sqrt{NT}(\tilde{\rho} - \rho) = O_p(1)$. Then, we can show, as in Park and Simar (1994) or Park, Sickles and Simar (1998), that

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

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 $\tilde{\beta}$, the GLS within $\tilde{\beta}_{GLS}$ and our two efficient estimators $\hat{\beta}$, $\hat{\beta}_{GLS}$. The finite sample performances are compared through the following Monte-Carlo (MC) scenarios.

We simulated for *Model 1* samples of size $N = 20, 100, 1000$ with $T = 12, 60$ in a model with $d = 2$ regressors. In each MC sample, the regressors were generated according to a bivariate VAR model:

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

where $\sigma_X = 1$ and $R = \begin{pmatrix} 4 & 0.05 \\ 0.05 & 4 \end{pmatrix}$. Then the obtained values of X_{it} were shifted around three different means to obtain almost 3 balanced groups of firms from smaller to larger. We fixed $\mu_1 = (5 \quad 5)', \mu_2 = (7.5 \quad 7.5)', \mu_3 = (10 \quad 10)'$. The idea is to generate a reasonable 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 parts (the individual effects) were generated independently of the regressors as $B - Expo(\mu_\alpha)$ where we chose for the exponential distribution a mean $\mu_\alpha = 1$ and for the upper boundary a value of $B = 1$. Since the y are often measured in logarithms (like in Cobb Douglas production functions), this involves an average inefficiency score $E\left(\exp\{-Expo(\mu_\alpha)\}\right) = 0.50$. 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 where obtained from $M = 500$ 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. In Tables 1–4, we display the sum of the two MC mean squared 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 bandwidth s we selected an optimal fixed value 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^* which minimizes the MSE. In all the tried scenarios, the results were not very sensitive to the choice of s in the above grid. For the analysis of a real data set in Section 4, we will propose a data driven method based on a bootstrap algorithm.

N	T	$\tilde{\beta}$	$\tilde{\beta}_{GLS}$	$\hat{\beta}$	$\hat{\beta}_{GLS}$	s^*
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 multilplied by 10^4. Here $\rho = 0.7$, $\sigma = 0.5$ and $\mu_\alpha = 1$. For $N = 1000$ only $M = 100$ replications were performed.*

In the situation of Table 1, with $\rho = 0.7$, we see a clear improvement in the use of the efficient estimator, in particular, compared with the OLS within. Note also how the GLS within dominates the OLS. By looking at the two versions of our efficient estimator, it is not clear that we gain by choosing 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 estimate ρ better and the improvement is still better as confirmed in Table 2 with $T = 60$.

When the autocorrelation coefficient 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 investigated 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^*
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 multilplied by* 10^4. *Here* $\rho = 0.7$, $\sigma = 0.5$ *and* $\mu_\alpha = 1$. *For* $N = 1000$ *only* $M = 100$ *replications were performed.*

N	T	$\tilde{\beta}$	$\tilde{\beta}_{GLS}$	$\hat{\beta}$	$\hat{\beta}_{GLS}$	s^*
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 multilplied by* 10^4. *Here* $\rho = 0.1$, $\sigma = 0.5$ *and* $\mu_\alpha = 1$. *For* $N = 1000$ *only* $M = 100$ *replications were performed.*

N	T	$\tilde{\beta}$	$\tilde{\beta}_{GLS}$	$\hat{\beta}$	$\hat{\beta}_{GLS}$	s^*
20	60	3.622	3.535	3.498	3.498	0.9
100	60	0.7428	0.7222	0.7028	0.7028	0.3
1000	60	0.0697	0.0677	0.0679	0.0679	1.1

Table 4: *Monte-Carlo MSE of the estimators of* β *with* $M = 500$ *replications. The figures for the MSE are multilplied by* 10^4. *Here* $\rho = 0.1$, $\sigma = 0.5$ *and* $\mu_\alpha = 1$. *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 estimation of ρ.

N	T	$\tilde{\beta}$	$\tilde{\beta}_{GLS}$	$\hat{\beta}$	$\hat{\beta}_{GLS}$	s^*
20	12	20.36	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 β with $M = 500$ replications. The figures for the MSE are multilplied by 10^4. Here $\rho = 0$, $\sigma = 0.5$ and $\mu_\alpha = 1$.*

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 β for the different sample sizes analyzed here.

Once the efficient estimators of β are obtained, estimators of ρ, 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. To compute the efficient estimator $\hat{\beta}$ we need to choose a bandwidth s. In a real data framework, we propose the following bootstrap procedure.

For a given s we compute $\hat{\beta}(s)$ and the resulting improved estimators $\widehat{\alpha}_i(s)$, $\hat{\rho}(s)$ and $\hat{\sigma}^2(s)$. 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):

$$\begin{aligned} Y_{it}^* &= X_{it}'\hat{\beta}(s) + \widehat{\alpha}_i(s) + \varepsilon_{it}^* \quad ; \quad i = 1, \ldots, N; t = 1, \ldots, T \\ \varepsilon_{it}^* &= \hat{\rho}(s)\varepsilon_{i,t-1}^* + u_{it}^*, \end{aligned} \tag{11}$$

where $u_{it}^* \sim \mathcal{N}(0, \hat{\sigma}^2(s))$. Each pseudo-sample provides $\hat{\beta}^{*b}(s)$. We compute the criterion value:

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

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

4 Empirical Illustration: Deregulatory Dynamics in the U.S. Banking Industry: 1984-1996

Since the early 1980's U. S. 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 1990's 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 U.S. banking environment. 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 1980's 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. For comprehensive discussions of these deregulatory issues and the industry's reactions and adjustments to them see Berger et al. (1995) and Humphrey and Pulley (1997).

Previous studies of banking productivity and efficiency have relied on three basic methods for productivity and efficiency measurement: linear programming, maximum likelihood, and ordinary least squares or instrumental variable estimation. Berger and Humphrey (1996) provide a general description of these methods. Our focus here is on efficient and robust measurement of productivity and efficiency in a setting in which the regulatory climate has been steadily altered, forcing firms to adjust to a best practice technology using resource allocations that are increasingly unconstrained by financial regulation.

The data set consists of about 2,550 U. S. banks from the first quarter 1984 through the fourth quarter 1996. We divide the data set into three subsamples based on differing regulatory environments: statewide branching (State), limited branching (Limit), and no branching (Unit). These samples contain 700, 1,100, and 750 banks each. Others have further separated the data according to bank size (Akhavein et al., 1997). The production and cost data 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 U.S. banking data. The 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 U. S. banking industry. Data on over one hundred variables was collected from the Call Reports and the FDIC Summary of Deposits.

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 U. S. $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 U. S. 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.

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). 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).

Outputs, inputs and price definitions used in this paper are consistent with those used in previous studies (Berger, 1991, and Berger et al., 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

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)[1].

We model the multiple output/multiple input banking technology using the output distance function. The output distance function, $D(Y, X) \leq 1$, provides a radial measure of technical efficiency by specifying the fraction of aggregated outputs (Y) produced given chosen inputs (X). An m-output, n-input deterministic distance function can be approximated by

[1] For a more detailed discussion of data and of structural modeling of competition in the U.S. banking industry, see Jayasiriya (2000).

$$\frac{\prod_j^m Y_j^{\gamma_j}}{\prod_k^n X_k^{\beta_k}} \leq 1$$

Here the $\gamma_j' s$ and the $\beta_k' s$ are weights describing the technology. Efficient production means that the index of total output cannot be increased without a reduction in a particular output or by an increase in a particular input, and thus the distance function is unity. The Cobb-Douglas stochastic distance frontier that we utilize below in our empirical illustration is derived by simply multiplying through by the denominator, approximating the terms using natural logarithms of output and inputs, and specifying an AR(1) additive error disturbance ε_{it} and an independent nonnegative stochastic term α_i for the firm specific level of radial technological inefficiency. The Cobb-Douglas stochastic distance frontier is thus

$$0 = \sum_j \gamma_j \ln y_{j,it} - \sum_k \beta_k \ln x_{k,it} + \alpha_i + \varepsilon_{it}$$

The output distance function is linearly homogenous in outputs (Diewert, 1982). We normalize the outputs with respect to the first output (real estate loans) and move (the negative log of) that output to the left-hand side. This provides us with a panel model that conforms our basic model (1). Since firm-specific technical efficiency terms are radial measures, it should be levels, not ratios that require augmentation in order for the firm to be technically efficient. Thus in order to interpret the effects as radial efficiency measures, the right-hand-side output mix is assumed to be exogeneous and thus uncorrelated with the firm effects and with the stochastic error. This is in keeping with treatments in the efficiency and productivity literature (Fixler and Zieschang, 1992, Lovell et al., 1994, and Grosskopf et al., 1997).

Table 6 provides variable definitions and summary statistics for the output and input data used in our distance function analysis for the three regulatory environments. We add to the analysis quarterly dummy variables and a time trend to identify overall technical change in the banking industry.

Results for the within OLS $\tilde{\beta}$, our semiparametric efficient estimators $\hat{\beta}$ and $\hat{\beta}_{GLS}$ for Model 1 are provided in Tables 7–9 for State, Limit, and Unit banks. The numbers between parenthesis are the corresponding estimated standard deviations (multiplied by 10^3). Bandwidth selection is based on the bootstrap procedure outlined in Section 3 and yielded choices for s^* equal to 1.7, 1.9, and 1.3. Serial correlation estimates range between 0.8838

Variable	Definition	State	Limit	Unit
reln	Log of real estate loans	9.974	9.334	8.969
ciln	Log of commercial and industrial loans	9.029	8.124	8.063
inln	Log of installment loans	8.638	8.366	7.900
CD	Log of certificate of deposits	8.722	8.078	7.700
DD	Log of demand deposits	9.412	8.524	8.395
OD	Log of retail time and savings deposits	10.836	10.365	10.115
lab	Log of labor	5.892	5.088	4.928
cap	Log of capital	7.131	6.402	6.181
purf	Log of purchased funds	11.220	10.689	10.478
	Number of observations	36400	57200	39000

Table 6: *Data Description-Sample Means.*

and 0.9428. As expected from our simulation study, for such large sample sizes, our two semiparametric efficient estimators are very similar. So, in all the following comments we used the semiparametric efficient estimators $\hat{\beta}$ for the comparison of the three regulatory regimes.

Results for the respective estimators across the three regulatory regimes do not indicate any significant scale economies while technological change correspondingly averaged between 1% and 3 %/year.

Variable	Within OLS $\tilde{\beta}$	Semipar-efficient $\hat{\beta}$	Semipar-efficient $\hat{\beta}_{GLS}$
ciln	0.3169	0.2542 (2.811)	0.2506 (2.854)
inln	0.1856	0.2211 (3.011)	0.2223 (3.058)
CD	-0.0861	-0.0216 (2.923)	-0.0192 (2.968)
DD	-0.0890	-0.0349 (4.731)	-0.0323 (4.804)
OD	-0.1231	-0.0781 (1.937)	-0.0734 (1.967)
lab	-0.2481	-0.1532 (6.832)	-0.1412 (6.937)
cap	-0.0098	-0.0677 (4.590)	-0.0698 (4.660)
purf	-0.4491	-0.3536 (7.640)	-0.3357 (7.757)
time	-0.0013	-0.0063 (0.298)	-0.0071 (0.303)
q1	0.0097	0.0081 (1.291)	0.0081 (1.311)
q2	-0.0061	-0.0025 (1.435)	-0.0021 (1.457)
q3	-0.0053	-0.0031 (1.213)	-0.0025 (1.321)
Adjusted R^2	0.8782	0.8344	0.8263

Table 7: *Results for STATE BANKS. The numbers between parenthesis are the corresponding estimated standard deviations (multiplied by 10^3). Note that here $\widetilde{\rho} = 0.8838$, $\widehat{\rho} = 0.9229$, $\widetilde{\sigma} = 0.1232$ and $\widehat{\sigma} = 0.1251$.*

Variable	Within OLS $\tilde{\beta}$	Semipar-efficient $\hat{\beta}$	Semipar-efficient $\hat{\beta}_{GLS}$
ciln	0.1611	0.1227 (1.679)	0.1209 (1.669)
inln	0.2947	0.2561 (2.637)	0.2522 (2.621)
CD	-0.0265	-0.0022 (0.685)	-0.0018 (0.681)
DD	-0.0617	-0.0104 (2.959)	-0.0100 (2.942)
OD	-0.1328	-0.0720 (1.324)	-0.0658 (1.317)
lab	-0.1925	-0.0787 (3.664)	-0.0699 (3.642)
cap	-0.0620	-0.0781 (3.043)	-0.0738 (3.025)
purf	-0.5655	-0.3256 (5.474)	-0.2980 (5.442)
time	-0.0038	-0.0087 (0.184)	-0.0092 (0.183)
q1	0.0060	0.0076 (0.800)	0.0075 (0.795)
q2	-0.0127	-0.0050 (0.901)	-0.0046 (0.896)
q3	-0.0241	-0.0126 (0.778)	-0.0118 (0.774)
Adjusted R^2	0.8280	0.7566	0.7438

Table 8: *Results for LIMIT BANKS. The numbers between parenthesis are the corresponding estimated standard deviations (multiplied by* 10^3*). Note that here* $\widetilde{\rho} = 0.8971$*,* $\widehat{\rho} = 0.9428$*,* $\widetilde{\sigma} = 0.0960$ *and* $\widehat{\sigma} = 0.0955$*.*

Variable	Within OLS $\tilde{\beta}$	Semipar-efficient $\hat{\beta}$	Semipar-efficient $\hat{\beta}_{GLS}$
ciln	0.2649	0.2037 (2.403)	0.2017 (2.386)
inln	0.2200	0.2554 (2.877)	0.2569 (2.858)
CD	-0.0437	0.0023 (2.221)	0.0033 (2.206)
DD	-0.0427	-0.0073 (3.834)	-0.0069 (3.809)
OD	-0.1000	-0.0769 (1.733)	-0.0733 (1.721)
lab	-0.2600	-0.0994 (5.013)	-0.0923 (4.980)
cap	-0.0321	-0.0664 (3.720)	-0.0658 (3.695)
purf	-0.5015	-0.3712 (7.356)	-0.3532 (7.307)
time	-0.0034	-0.0067 (0.235)	-0.0070 (0.234)
q1	0.0009	0.0053 (1.120)	0.0055 (1.112)
q2	-0.0126	-0.0066 (1.245)	-0.0063 (1.237)
q3	-0.0170	-0.0101 (1.071)	-0.0098 (1.064)
Adjusted R^2	0.8367	0.7956	0.7901

Table 9: *Results for UNIT BANKS. The numbers between parenthesis are the corresponding estimated standard deviations (multiplied by* 10^3*). Note that here* $\widetilde{\rho} = 0.8842$*,* $\widehat{\rho} = 0.9200$*,* $\widetilde{\sigma} = 0.1085$ *and* $\widehat{\sigma} = 0.1077$*.*

Relative efficiencies were highest for unit banks, followed by limit, and then state banks (Table 10). Their distribution is given in Figure 1. We perform a Kolmogorov-Smirnoff test of the within residuals against normality. Figures 2-4 show plots of the distribution functions for the standardized normal and the standardized within residuals. Clearly, the two distributions correspond quite closely. However, because sample sizes are 39,000, 57,2000, and 36,400 for the unit, limit, and state bank samples, the formal test rejects normality at the 5% level. We can also examine the differences in efficiencies before and after the passing of the Reigle-Neal Act in the early 1990's by estimating separate regressions for the pre and post 1990 observations. This analysis indicates that relative efficiencies rose 2%, 35%, and 13% for unit, limit, and state banks, averaging an almost 20% gain in productive efficiency between the 1980's and the 1990's.

Banks	Average $\hat{\alpha}_i$
State	0.3869 (0.1730)
Limit	0.3923 (0.1821)
Unit	0.4286 (0.1910)

Table 10: *Average Relative Efficiencies for State, Limit and Unit Banks.*

5 Conclusion

This paper has introduced a new semiparametric efficient estimator for random effects panel data models with AR(1) errors. We have considered specifications in which regressors and effects are uncorrelated (Model 1) and in which they are correlated (Model 2). We have also considered estimation and inference in boundary function models in which panel random effect estimators have been used widely in the empirical productivity literature. Analytical results were supplemented with Monte Carlo simulations and with an empirical illustration of banking efficiency that finds a 20% increase in banking technical efficiency between the 1980's and the 1990's.

Figure 1. Relative Efficiencies in Semiparametric Efficient Regression

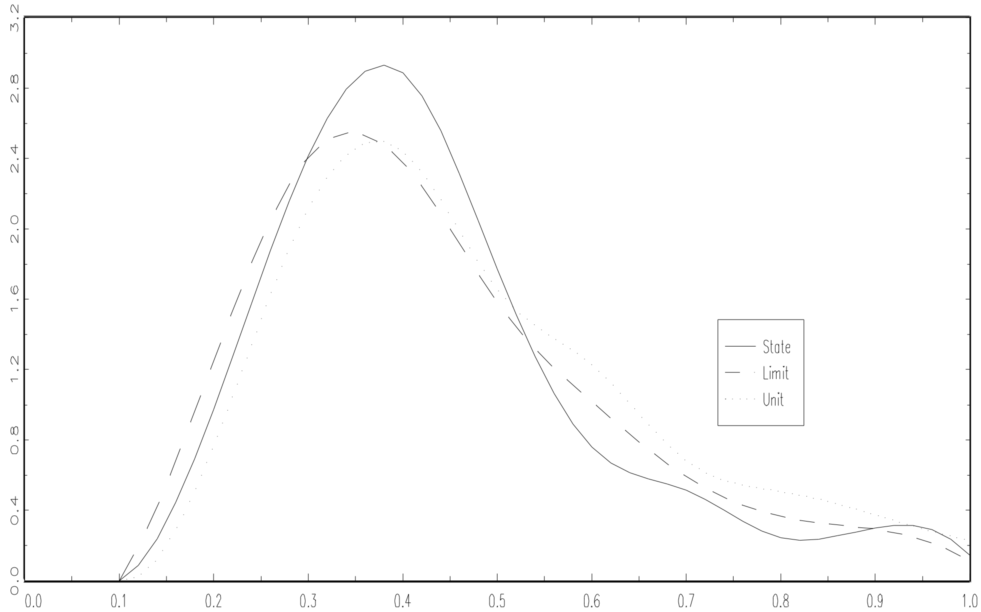


Figure 2. The Distribution of the Standardized Within Residuals of State Banks

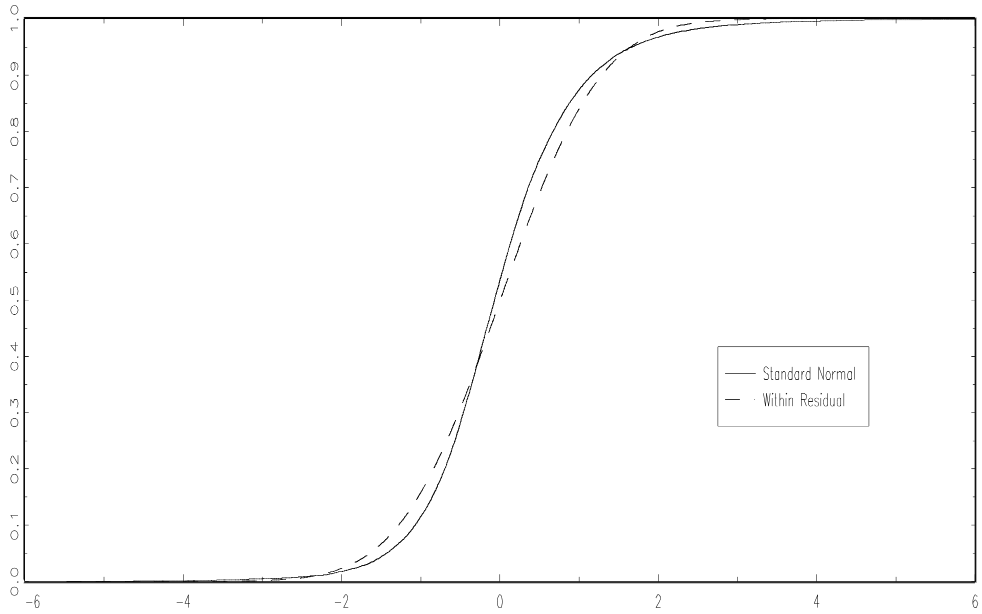


Figure 3. The Distribution of the Standardized Within Residuals of Limit Banks

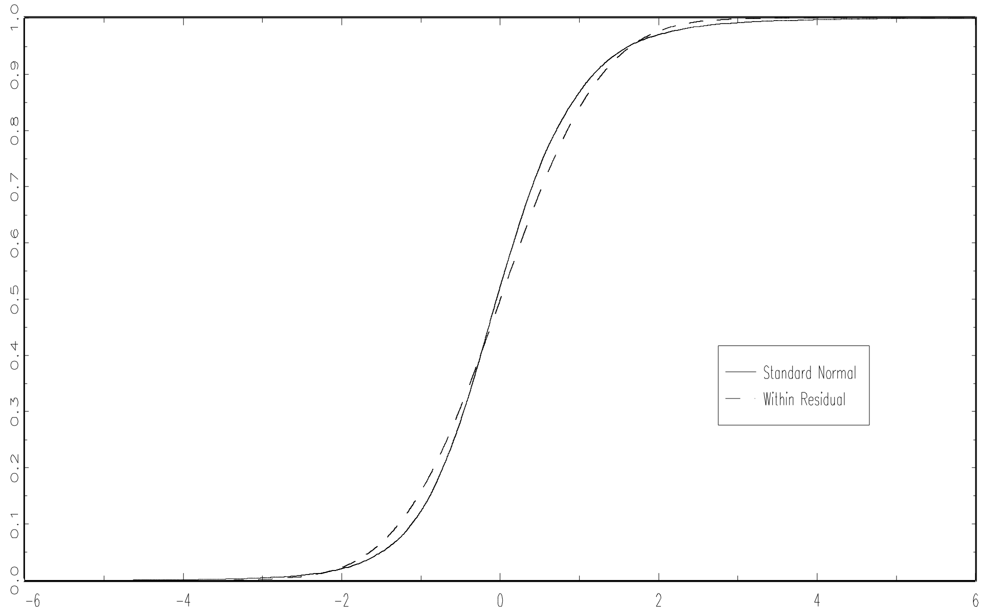
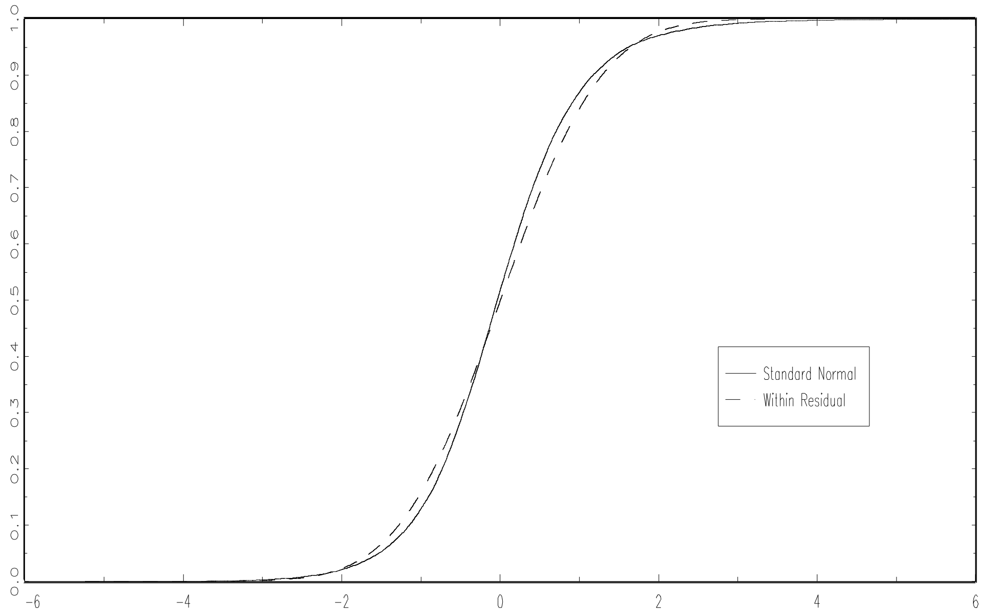


Figure 4. The Distribution of the Standardized Within Residuals of Unit Banks



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 β 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 $\{P_{(\beta,\eta)} : \beta \in \mathbb{R}^d, \eta \in S \text{ open} \subset \mathbb{R}^k\}$ and at every (β_0, η_0) the mapping $(\beta, \eta) \to P_{(\beta,\eta)}$ 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 $(\beta, \rho, \sigma^2, h)$. Let $(\beta_0, \rho_0, \sigma_0^2, h_0)$ denote the true value of the true underlying P. Write η_1, η_2 for ρ, σ^2 with $\eta_{10} = \rho_0, \eta_{20} = \sigma_0^2$. Consider a class of functions $h_{\eta_3}(\cdot, \cdot)$ indexed by $\eta_3 \in \mathbb{R}^1$ where h_{η_3} is identical to the true h_0 when $\eta_3 = 0$, for example, one may take $h_{\eta_3}(\cdot, \cdot) = h_0(\cdot - \eta_3, \cdot)$. Let $P_{(\beta,\eta_1,\eta_2,\eta_3)}$ denote a distribution characterized by $(\beta, \eta_1, \eta_2, \eta_3)$. This means that the true P can be written as $P_{(\beta_0,\eta_{10},\eta_{20},0)}$. Then, the class of probability distribution $\{P_{(\beta,\eta_1,\eta_2,\eta_3)} : \beta \in \mathbb{R}^d, -1 < \eta_1 < 1, \eta_2 > 0, \eta_3 \in \mathbb{R}^1\}$ is a submodel of $\mathbb{P}$ passing through the true P and is regular if h_{η_3}, as functional of η_3, is "smooth" in a certain sense (Ibragimov and Has'minskii, 1981).

Suppose $P(= P_{(\beta_0,\eta_0)})$ belongs to a regular parametric submodel $\mathbb{P}_0$ of $\mathbb{P}$. Then the notion of information bound and efficient estimation of β are well defined. Let $L(X, Y, \beta, \eta)$ denote the log likelihood of an observation from $P_{(\beta,\eta)}$ and let $\ell_\beta(X, Y) = \partial L/\partial\beta|_{(\beta_0,\eta_0)}$ and $\ell_{\eta_j}(X, Y) = \partial L/\partial\eta_j|_{(\beta_0,\eta_0)}$ where $\eta = (\eta_1, \ldots, \eta_k)$. Then,

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

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

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

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

Moreover, it can be also written as

$$I(P; \beta, \mathbb{P}_0) = E\ell^* \ell^{*'}$$

where

$$\ell^* = \ell_\beta - \pi(\ell_\beta | [\ell_\eta]),$$

$[\ell_\eta]$ denotes the linear span generated by $\{\ell_{\eta_j}\}_{j=1}^{k}$, and $\pi(\ell|S)$ denotes the vector of projections of each component of ℓ 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 β is called *efficient* if it is asymptotically normal with mean zero and variance $N^{-1}I^{-1}(P;\beta,\mathbb{P}_0)$.

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 $I(P;\beta,\mathbb{P}) = \inf\{I(P;\beta,\mathbb{P}_0) : \mathbb{P}_0$ is a regular parametric submodel of $\mathbb{P},\ P \in \mathbb{P}_0\}$ be the information bound for estimating β under $\mathbb{P}$. An estimator $\hat\beta_N$ is now called *efficient* in $\mathbb{P}$ if

$$\mathcal{L}_P(\sqrt{N}(\hat\beta_N - \beta)) \to \mathcal{N}(0, I^{-1}(P;\beta,\mathbb{P})).$$

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

$$I(P;\beta,\mathbb{P}) = E\ell^*\ell^{*'}(X,Y)$$

where

$$\ell^* = \ell_\beta - \pi(\ell_\beta|V)$$

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

B Lemmas and Proofs

B.1 Proof of Theorem 2.1.

With the notations of Section 2.1, the log-likelihood of (X,Y) can be written as

$$\begin{aligned} L(X,Y;\beta,\rho,\sigma^2,h,g) &= \log g(X) - \frac{T}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(1-\rho^2) \\ &\quad -\frac{1}{2\sigma^2}\left\{(1-\rho^2)(Y_1 - X_1'\beta)^2 + \sum_{t=2}^{T}(Y_t - X_t'\beta - \rho(Y_{t-1} - X_{t-1}'\beta))^2\right\} \\ &\quad +\frac{W^2}{2v^2} + \log f(W) + \frac{1}{2}\log(2\pi v^2). \end{aligned}$$

Then, the score functions are given by

$$\begin{aligned}
\ell_\beta &:= \frac{\partial L}{\partial \beta} \\
&= \frac{1}{\sigma^2}\left\{(1-\rho^2)Z_1X_1 + \sum_{t=2}^{T}(Z_t-\rho Z_{t-1})(X_t-\rho X_{t-1})\right\} - \frac{f^{(1)}}{f}(W)\widetilde{X} \qquad \text{(B.1)} \\
\ell_{\sigma^2} &:= \frac{\partial L}{\partial \sigma^2} \\
&= \frac{1}{2\sigma^2}\left\{\frac{Z_1^2}{\sigma^2/(1-\rho^2)} + \sum_{t=2}^{T}(Z_t-\rho Z_{t-1})^2/\sigma^2 \right.\\
&\qquad \left. + \frac{\int \phi_v(W-u)[(W-u)^2/v^2 - T]h(u)du}{f(W)}\right\}. \qquad \text{(B.2)}
\end{aligned}$$

For computing $\ell_\rho := \partial L/\partial \rho$ we need some ingredients and we introduce more notations. Let $k(\rho) = 2\rho + 2(T-1)(1-\rho)$ and $S = S(\rho,\beta) := \sum_{t=1}^{T} d_t(Y_t - X_t'\beta)$ where $d_t = d_t(\rho)$ are another weights defined by

$$d_t = d_t(\rho) = \begin{cases} 1/k(\rho) & , \text{ for } t = 1 \text{ or } T \\ 2(1-\rho)/k(\rho) & , \text{ for } t = 2,\dots,T-1. \end{cases}$$

Note that $\sum_{t=1}^{T} d_t = 1$. Then it follows that

$$\begin{aligned}
\frac{\partial}{\partial \rho}(v^2) &= \frac{k(\rho)\sigma^2}{T(\rho)^2} \\
\frac{\partial}{\partial \rho}(W) &= \frac{k(\rho)(W-S)}{T(\rho)} \\
\frac{\partial}{\partial \rho}(W^2/2v^2) &= k(\rho)W(W-2S)/2\sigma^2.
\end{aligned}$$

Thus, with these preliminary results we can obtain:

$$\begin{aligned}
\ell_\rho &= -\frac{\rho}{1-\rho^2} + \frac{1}{2\sigma^2}\left\{2\rho(Y_1 - X_1'\beta)^2 + \sum_{t=2}^{T} 2(Y_t - X_t'\beta - \rho(Y_{t-1} - X_{t-1}'\beta)) \right.\\
&\qquad \left. \cdot (Y_{t-1} - X_{t-1}'\beta) + \frac{\int \phi_v(W-u)k(\rho)(u^2 - 2uS)h(u)du}{f(W)}\right\}. \qquad \text{(B.3)}
\end{aligned}$$

B.1.1 Computation of the efficient score ℓ^* for estimating β. Let V be the tangent space for the nuisance parameters σ^2, ρ, g, h. Then, $V = V_1 + V_2 + V_3 + V_4$ where $V_1 = [\ell_{\sigma^2}]$, $\quad V_2 = [\ell_\rho]$ and $V_3 = \{a(X) \,:\, a \in L_2(P), Ea(X) = 0\}, V_4 = \{b(W) \,:\, b \in L_2(P), Eb(W) = 0\}$. Here and below $[\ell]$ means the linear space spanned by ℓ. We know that the efficient score ℓ^* is given by $\ell^* = \ell_\beta - \pi(\ell_\beta|V)$. The computation of ℓ^* will be done going through the following six steps (i)–(vi).

Step (i): ℓ_β is perpendicular to V_3 in the $L_2(P)$ sense.

Proof. We need to show that $E(\ell_\beta|X) = 0$. This follows since X is independent of W and $Z_1, \ldots, Z_T$, and thus for $t = 1, \ldots, T$

$$E(Z_t|X) = E(Z_t) = E(\varepsilon_t - \sum_{s=1}^{T} c_s \varepsilon_s) = 0$$
$$E\left\{\frac{f^{(1)}}{f}(W)|X\right\} = E\left\{\frac{f^{(1)}}{f}(W)\right\} = \int f^{(1)}(w)dw = 0. \qquad \blacksquare$$

From Step (i), ℓ^* is now given by

$$\begin{aligned}\ell^* &= \ell_\beta - \pi(\ell_\beta|V_1 + V_2 + V_4) \\ &= \ell_\beta - \pi(\ell_\beta|V_4) - \pi(\ell_\beta - \pi(\ell_\beta|V_4)|\widetilde{V}),\end{aligned} \tag{B.4}$$

where $\widetilde{V}$ is the linear span generated by $\ell_{\sigma^2} - \pi(\ell_{\sigma^2}|V_4)$ and $\ell_\rho - \pi(\ell_\rho|V_4)$, *i.e.*

$$\widetilde{V} := [\ell_{\sigma^2} - \pi(\ell_{\sigma^2}|V_4), \ell_\rho - \pi(\ell_\rho|V_4)].$$

Before moving to Step (ii), we need the following lemma.

Lemma B.1 *W is independent of $Z_1, \ldots, Z_T$.*

Proof. From the definition of W in Section 2.1, we can write $W = \alpha + \widetilde{\varepsilon}$ where $\widetilde{\varepsilon} = \sum_{t=1}^{T} c_t \varepsilon_t$. We observe that $Z_1 = \varepsilon_1 - \widetilde{\varepsilon}$ and $Z_t - \rho Z_{t-1} = u_t - (1-\rho)\widetilde{\varepsilon},\ t = 2, \ldots, T$. The joint pdf of $R_1 = \varepsilon_1 - \widetilde{\varepsilon}, R_2 = u_2 - (1-\rho)\widetilde{\varepsilon}, \ldots, R_{T-1} = u_{T-1} - (1-\rho)\widetilde{\varepsilon}$ and $R_T = \widetilde{\varepsilon}$ is given by

$$\begin{aligned}p(r_1, \ldots, r_T; \rho, \sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{T/2} (1-\rho^2)^{1/2} \{(1+\rho) + (T-1)(1-\rho)\} \\ &\quad \times \exp\Big[-\frac{1}{2\sigma^2}\Big\{(1-\rho^2)(r_1 + r_T)^2 + \sum_{t=2}^{T-1}(r_t + (1-\rho)r_T)^2 \\ &\qquad + ((1+\rho)r_1 + (r_2 + \ldots + r_{T-1}) - (1-\rho)r_T)^2\Big\}\Big].\end{aligned}$$

Here, in the exponent, all the cross products of $r_t,\ t = 1, \ldots, T-1$ with r_T disappear. In fact, the exponent part can be written as

$$(1-\rho^2)r_1^2 + \sum_{t=2}^{T-1} r_t^2 + \left\{(1+\rho)r_1 + \sum_{t=2}^{T-1} r_t\right\}^2 + T(\rho)r_T^2.$$

This means that $(R_1, \ldots, R_{T-1})$ is independent of $R_T = \widetilde{\varepsilon}$. Since $Z_T - \rho Z_{T-1} = u_T - (1-\rho)\widetilde{\varepsilon}$ is a linear combination of $R_1, \ldots, R_{T-1}$, it is independent of $\widetilde{\varepsilon}$ too. Thus, $\widetilde{\varepsilon}$ is independent of $Z_1, Z_2 - \rho Z_1, \ldots, Z_T - \rho Z_{T-1}$. Furthermore, since α is independent of $(\varepsilon_1, u_2, \ldots, u_T)$ too and $(Z_1, \ldots, Z_T)$ has a one-to-one correspondence with $(Z_1, Z_2 - \rho Z_1, \ldots, Z_T - \rho Z_{T-1})$, we conclude that W is independent of $Z_1, \ldots, Z_T$. ■

Step (ii): $\pi(\ell_\beta | V_4) = -\frac{f^{(1)}}{f}(W)\, E(\widetilde{X})$.

Proof. Lemma B.1 implies that $E(Z_t|W) = E(Z_t) = 0,\ t = 1, \ldots, T$. Since $\pi(\ell_\beta|V_4) = E(\ell_\beta|W)$ and $E(\widetilde{X}|W) = E(\widetilde{X})$, the result follows. ■

Now we have from Lemma B.1 and Step (ii)

$$\begin{aligned} \ell_\beta - \pi(\ell_\beta|V_4) &= \frac{1}{\sigma^2}\left\{(1-\rho^2)Z_1X_1 + \sum_{t=2}^{T}(Z_t - \rho Z_{t-1})(X_t - \rho X_{t-1})\right\} \\ &\quad -\frac{f^{(1)}}{f}(W)(\widetilde{X} - E\widetilde{X}). \end{aligned} \tag{B.5}$$

Step (iii): $\ell_\beta - \pi(\ell_\beta|V_4)$ is perpendicular to $\ell_{\sigma^2} - \pi(\ell_{\sigma^2}|V_4)$.

Proof. $\ell_\beta - \pi(\ell_\beta|V_4)$ is already perpendicular to $\pi(\ell_{\sigma^2}|V_4)$, hence we need to show that it is perpendicular to ℓ_{σ^2}. But this is true since $Z_t \stackrel{\mathrm{d}}{=} -Z_t$ and $X, W, (Z_1, \ldots, Z_T)$ are independent.■

The following lemma will be useful for our Step (iv).

Lemma B.2 *$E(S|W) = W$.*

Proof. We have

$$\begin{aligned} E(S|W) &= E(S - W|W) + W \\ &= E(\sum_{t=1}^{T} d_t Z_t|W) + W \\ &= W. \quad ■ \end{aligned}$$

Step (iv): The projection of ℓ_ρ on V_4 is given by

$$\begin{aligned} \pi(\ell_\rho|V_4) &= -\frac{\rho}{1-\rho^2} + \frac{1}{2\sigma^2}k(\rho)W^2 - \frac{(T-1)(1-\rho)^2}{(1-\rho^2)T(\rho)} \\ &\quad + \frac{1}{2\sigma^2}\frac{\int \phi_v(W-u)k(\rho)(u^2 - 2uW)h(u)du}{f(W)} \end{aligned}$$

Proof. Since $Y_t - X_t'\beta = Z_t + W$, we observe that

$$\begin{aligned} &2\rho(Y_1 - X_1'\beta)^2 + \sum_{t=2}^{T} 2(Y_t - X_t'\beta - \rho(Y_{t-1} - X_{t-1}'\beta))(Y_{t-1} - X_{t-1}'\beta) \\ &= k(\rho)W^2 + 2\rho Z_1^2 + 2\sum_{t=2}^{T}(Z_t - \rho Z_{t-1})Z_{t-1} + 2k(\rho)W\sum_{t=1}^{T} d_t Z_t. \end{aligned}$$

Noting that

$$E(Z_1^2|W) = E(Z_1^2) = \frac{(T-1)(1-\rho)^2}{(1-\rho^2)T(\rho)}\sigma^2, \quad \text{and} \tag{B.6}$$

$$E\left[(Z_t - \rho Z_{t-1})Z_{t-1}|W\right] = E\left[(Z_t - \rho Z_{t-1})Z_{t-1}\right] = -\frac{1-\rho}{T(\rho)}\sigma^2, \tag{B.7}$$

we can compute $\pi(\ell_\rho|V_4) = E(\ell_\rho|W)$ from (B.3) by using Lemma B.2. This provides the desired results. ■

Step (v): It follows that

$$\ell_\rho - \pi(\ell_\rho|V_4) = \frac{1}{\sigma^2}\left\{\rho Z_1^2 + \sum_{t=2}^{T}(Z_t - \rho Z_{t-1})Z_{t-1}\right\} - \frac{k(\rho)}{T(\rho)}\,\frac{f^{(1)}}{f}(W)\sum_{t=1}^{T} d_t Z_t + \frac{(T-1)(1-\rho)^2}{(1-\rho^2)T(\rho)}$$

Proof. Note that $S - W = \sum_{t=1}^{T} d_t Z_t$. Thus,

$$\begin{aligned}\ell_\rho - \pi(\ell_\rho|V_4) \;&=\; \frac{1}{2\sigma^2}\left\{2\rho Z_1^2 + 2\sum_{t=2}^{T}(Z_t - \rho Z_{t-1})Z_{t-1} + 2k(\rho)W(S-W)\right\}\\ &\quad + \frac{(T-1)(1-\rho)^2}{(1-\rho^2)T(\rho)} + \frac{1}{2\sigma^2}\frac{\int 2\phi_v(W-u)uh(u)\,du\cdot k(\rho)(W-S)}{f(W)}.\end{aligned}$$

Since $\int \phi_v(W-u)uh(u)\,du = v^2 f^{(1)}(W) + Wf(W)$ and $v^2/\sigma^2 = 1/T(\rho)$, the last term in the above expression is equal to

$$\frac{1}{\sigma^2}k(\rho)W(W-S) + \frac{k(\rho)}{T(\rho)}\,\frac{f^{(1)}}{f}(W)(W-S).$$

This concludes the proof. ■

Step (vi): $\ell_\beta - \pi(\ell_\beta|V_4)$ is perpendicular to $\ell_\rho - \pi(\ell_\rho|V_4)$.

Proof. This is obvious, since $Z_t \overset{\mathrm{d}}{=} -Z_t$, $E[\frac{f^{(1)}}{f}(W)] = 0$, and $X, W, (Z_1, \ldots, Z_T)$ are independent. ■

Since now $\ell_\beta - \pi(\ell_\beta|V_4)$ is perpendicular to $\widetilde{V}$, we have from (B.4) and (B.5)

$$\begin{aligned}\ell^* \;&=\; \frac{1}{\sigma^2}\left\{(1-\rho^2)Z_1X_1 + \sum_{t=2}^{T}(Z_t - \rho Z_{t-1})(X_t - \rho X_{t-1})\right\}\\ &\quad - \frac{f^{(1)}}{f}(W)(\widetilde{X} - E\widetilde{X}).\end{aligned}$$

B.1.2 Computation of the information bound I for estimating β. Observe that

$$
\begin{aligned}
&(1-\rho^2)Z_1X_1 + \sum_{t=2}^{T}(Z_t-\rho Z_{t-1})(X_t-\rho X_{t-1}) \\
&= (1-\rho^2)Z_1(X_1-\widetilde{X}) + \sum_{t=2}^{T}(Z_t-\rho Z_{t-1})\{X_t-\widetilde{X}-\rho(X_{t-1}-\widetilde{X})\} \\
&\quad +\widetilde{X}\left\{(1-\rho^2)Z_1 + (1-\rho)\sum_{t=2}^{T}(Z_t-\rho Z_{t-1})\right\}.
\end{aligned}
$$

The second term vanishes since $(1-\rho^2)Z_1+(1-\rho)\sum_{t=2}^{T}(Z_t-\rho Z_{t-1}) = \sum_{t=1}^{T} c_tZ_t = 0$. Thus,

$$
\begin{aligned}
\ell^* &= \frac{1}{\sigma^2}\left\{(1-\rho^2)Z_1(X_1-\widetilde{X}) + \sum_{t=2}^{T}(Z_t-\rho Z_{t-1})\{X_t-\widetilde{X}-\rho(X_{t-1}-\widetilde{X})\}\right\} \\
&\quad -\frac{f^{(1)}}{f}(W)(\widetilde{X}-E\widetilde{X}).
\end{aligned}
$$

Using (B.6) and (B.7), a straightforward calculation shows that

$$
\begin{aligned}
&E(Z_s-\rho Z_{s-1})(Z_t-\rho Z_{t-1}) = \begin{cases} \left[1-\dfrac{(1-\rho)^2}{T(\rho)}\right]\sigma^2 & \text{when } s=t \\ -\dfrac{(1-\rho)^2}{T(\rho)}\sigma^2 & \text{when } s\neq t \end{cases} \\
&E(Z_t-\rho Z_{t-1})Z_1 = -\frac{(1-\rho)}{T(\rho)}\sigma^2.
\end{aligned}
$$

Plugging these formulas in $E\ell^*\ell^{*\prime}$ yields

$$
\begin{aligned}
I &= \frac{1}{\sigma^2}\Big\{(1-\rho^2)E(X_1-\widetilde{X})(X_1-\widetilde{X})' \\
&\quad +\sum_{t=2}^{T}E(X_t-\widetilde{X}-\rho(X_{t-1}-\widetilde{X}))(X_t-\widetilde{X}-\rho(X_{t-1}-\widetilde{X}))' \\
&\quad -\frac{1}{T(\rho)}E[(1-\rho^2)(X_1-\widetilde{X})+(1-\rho)\sum_{t=2}^{T}E(X_t-\widetilde{X}-\rho(X_{t-1}-\widetilde{X}))] \\
&\qquad \cdot[(1-\rho^2)(X_1-\widetilde{X})+(1-\rho)\sum_{t=2}^{T}E(X_t-\widetilde{X}-\rho(X_{t-1}-\widetilde{X}))]'\Big\} \\
&\quad +E\{\frac{f^{(1)}}{f}(W)\}^2\, E(\widetilde{X}-E\widetilde{X})(\widetilde{X}-E\widetilde{X})'.
\end{aligned}
$$

The results then follows since

$$
(1-\rho^2)(X_1-\widetilde{X})+(1-\rho)\sum_{t=2}^{T}E(X_t-\widetilde{X}-\rho(X_{t-1}-\widetilde{X})) = \sum_{t=1}^{T}c_t(X_t-\widetilde{X}) = 0.
$$

B.2 Proof of Theorem 2.2

Define $S_i(\rho, \beta)$ in the same way as we defined, for instance $W_i(\rho, \beta)$, in Section 2.1, i.e.

$$S_i(\rho, \beta) = \sum_{t=1}^{T} d_t(\rho)(Y_{it} - X_{it}'\beta).$$

Below, we sometimes write $W_i, S_i, Z_{it}, \widetilde{X}_i$ and $\widetilde{X}.$ for $W_i(\rho, \beta), S_i(\rho, \beta), Z_{it}(\rho, \beta), \widetilde{X}_i(\rho)$ and $\widetilde{X}.(\rho)$ respectively when ρ and β are the true parameters.

First define

$$\begin{aligned} f_N(w) = f_N(w;\rho) \;&:=\; E\hat{f}_N(w;\rho,\beta) \\ &=\; K_s * f(w,\rho) + c \end{aligned}$$

Note that f_N depends only on ρ, not on β. Also define

$$\begin{aligned} I_{f,N} &:= \int \{\frac{f_N^{(1)}}{f_N}(w)\}^2 f(w)dw, \\ I_N &:= \sigma^{-2}\Sigma_1 + I_{f,N}\Sigma_2. \end{aligned}$$

The following results are standard in semiparametric literature: as $N \to \infty$ we have

$$E\{\frac{f_N^{(1)}}{f_N}(W_1) - \frac{f^{(1)}}{f}(W_1)\}^2 \longrightarrow 0 \tag{B.8}$$

$$E|\frac{f_N^{(2)}}{f_N}(W_1) - \frac{f^{(2)}}{f}(W_1)| \longrightarrow 0 \tag{B.9}$$

See, for example, Bickel and Ritov (1987) or Park, Sickles and Simar (1998) for the proofs of these results. The result (B.8) enables us to show

$$I_N \longrightarrow I \tag{B.10}$$

$$\{\widetilde{X}. - E(\widetilde{X}_1)\}N^{-1/2}\sum_{i=1}^{N}\frac{f_N^{(1)}}{f_N}(W_i) \xrightarrow{\text{p}} 0 \tag{B.11}$$

$$N^{-1/2}\sum_{i=1}^{N}(\widetilde{X}_i - E(\widetilde{X}_1))\{\frac{f_N^{(1)}}{f_N}(W_i) - \frac{f^{(1)}}{f}(W_i)\} \xrightarrow{\text{p}} 0. \tag{B.12}$$

First, (B.10) follows since

$$\begin{aligned} I_{f,N} - I_f \;&=\; E\{\frac{f_N^{(1)}}{f_N}(W_1) - \frac{f^{(1)}}{f}(W_1)\}^2 + 2E\{\frac{f_N^{(1)}}{f_N}(W_1) - \frac{f^{(1)}}{f}(W_1)\}\frac{f^{(1)}}{f}(W_1) \\ &\leq\; E\{\frac{f_N^{(1)}}{f_N}(W_1) - \frac{f^{(1)}}{f}(W_1)\}^2 + 2\Big\{E\Big(\frac{f_N^{(1)}}{f_N}(W_1) - \frac{f^{(1)}}{f}(W_1)\Big)^2 \cdot I_f\Big\}^{1/2} \\ &\longrightarrow 0. \end{aligned}$$

The result (B.11) follows since the left hand side has mean zero and variance given by

$$\frac{1}{N}\mathrm{var}(\widetilde{X}_1)\mathrm{var}\{\frac{f_N^{(1)}}{f_N}(W_1)\} \le \frac{1}{N}\mathrm{var}(\widetilde{X}_1)I_{f,N}.$$

Finally, (B.12) follows since the left hand side has mean zero and variance given by

$$\mathrm{var}(\widetilde{X}_1)E\{\frac{f_N^{(1)}}{f_N}(W_1) - \frac{f^{(1)}}{f}(W_1)\}^2.$$

Now, define

$$\begin{aligned} Q_N(r,b) \;:=\; & N^{-1/2}\sum_{i=1}^{N}\{\sigma^{-2}(1-r^2)Z_{i1}(r,b)X_{i1} \\ & +\sigma^{-2}\sum_{t=2}^{T}(Z_{it}(r,b) - rZ_{i,t-1}(r,b))(X_{it} - rX_{i,t-1}) \\ & -(\widetilde{X}_i(r) - \widetilde{X}.(r))\frac{f_N^{(1)}}{f_N}(W_i(r,b;r))\}. \end{aligned}$$

Note that if we replace $\tilde{\sigma}^2$ by σ^2 and $\hat{f}(\cdot;\tilde{\rho},\tilde{\beta})$ by $f_N(\cdot;\tilde{\rho})$ in the definition of $\hat{\beta}$, then we can write $\hat{\beta}$ as

$$\tilde{\beta} + N^{-1/2}\widehat{I}^{-1}Q_N(\tilde{\rho},\tilde{\beta}).$$

Also, note that by (B.11) and (B.12)

$$I_N^{-1}Q_N(\rho,\beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}). \tag{B.13}$$

Define

$$R_N(r,b) := N^{-1/2}\sum_{i=1}^{N}(\widetilde{X}_i(r) - \widetilde{X}.(r))\Big\{\frac{\hat{f}^{(1)}}{\hat{f}}(W_i(r,b);r,b) - \frac{f_N^{(1)}}{f_N}(W_i(r,b);r,b)\Big\}.$$

Then, the proof of the theorem is reduced to the proofs of

$$\widehat{I} - I_N \xrightarrow{\mathrm{p}} 0 \tag{B.14}$$

$$R_N(\tilde{\rho},\tilde{\beta}) \xrightarrow{\mathrm{p}} 0 \tag{B.15}$$

$$|Q_N(\tilde{\rho},\tilde{\beta}) - Q_N(\rho,\beta) - (\tilde{\rho}-\rho)\frac{\partial}{\partial\rho}Q_N(\rho,\beta) - (\tilde{\beta}-\beta)\frac{\partial}{\partial\beta}Q_N(\rho,\beta)| \xrightarrow{\mathrm{p}} 0 \tag{B.16}$$

$$N^{-1/2}\frac{\partial}{\partial\rho}Q_N(\rho,\beta) \xrightarrow{\mathrm{p}} 0 \tag{B.17}$$

$$N^{-1/2}\frac{\partial}{\partial\beta}Q_N(\rho,\beta) + I_N \xrightarrow{\mathrm{p}} 0. \tag{B.18}$$

The proofs of (B.14), (B.15) and (B.16) can be done similarly as in the proofs of Lemma A.2 and (A.16) of Park and Simar (1994). Hence they are omitted here. For (B.17), recall

$$\frac{\partial}{\partial\rho}W_i(\rho) = k(\rho)\{W_i(\rho) - S_i(\rho)\}/T(\rho).$$

We can write

$$
\begin{aligned}
& N^{-1/2}\frac{\partial}{\partial\rho}Q_N(\rho,\beta) \\
= \; & -N^{-1}\sigma^{-2}\sum_{i=1}^{N}\{2\rho Z_{i1}X_{i1}+\sum_{t=2}^{T}(Z_{it}-\rho Z_{i,t-1})X_{i,t-1}+k(\rho)(W_i-S_i)\widetilde{X}_i\} \\
& -N^{-1}k(\rho)T(\rho)^{-1}\sum_{i=1}^{N}\{\frac{f_N^{(2)}}{f_N}(W_i)-(\frac{f_N^{(1)}}{f_N})^2(W_i)\}(W_i-S_i)(\widetilde{X}_i-\widetilde{X}.) \\
& -N^{-1}\sum_{i=1}^{N}\frac{f_N^{(1)}}{f_N}(W_i)\frac{\partial}{\partial\rho}\Big(\widetilde{X}_i(\rho)-\widetilde{X}.(\rho)\Big).
\end{aligned}
$$

Observing that $E(S_i|W_i)=W_i$ and that $\frac{\partial}{\partial\rho}\widetilde{X}_i(\rho)$ is another weighted average of the X_{it}'s (this time the sum of the weights is zero) and so $E\{(\frac{\partial}{\partial\rho}(\widetilde{X}_i(\rho)-\widetilde{X}.(\rho))\}=0$, we can find that $N^{-1/2}\frac{\partial}{\partial\rho}Q_N(\rho,\beta)$ has mean zero and variance of order $1/N$. Thus (B.17) is established.

Finally, to prove (B.18), we find that

$$
\begin{aligned}
& N^{-1/2}\frac{\partial}{\partial\beta}Q_N(\rho,\beta) \\
= \; & -N^{-1}\sigma^{-2}\sum_{i=1}^{N}\Big\{(1-\rho^2)(X_{i1}-\widetilde{X}_i)(X_{i1}-\widetilde{X}_i)' \\
& \qquad +\sum_{t=2}^{T}(X_{it}-\widetilde{X}_i-\rho(X_{i,t-1}-\widetilde{X}_i))(X_{it}-\widetilde{X}_i-\rho(X_{i,t-1}-\widetilde{X}_i))'\Big\} \\
& +N^{-1}\sum_{i=1}^{N}\{\frac{f_N^{(2)}}{f_N}(W_i)-(\frac{f_N^{(1)}}{f_N})^2(W_i)\}(\widetilde{X}_i-\widetilde{X}.)\widetilde{X}_i'.
\end{aligned}
\tag{B.19}
$$

The first term of (B.19) tends to $-\sigma^{-2}\Sigma_1$ by WLLN. The second term has mean

$$
\begin{aligned}
& E\;\;\Big\{\frac{f_N^{(2)}}{f_N}(W_1)-(\frac{f_N^{(1)}}{f_N})^2(W_1)\Big\}E(\widetilde{X}_1-\widetilde{X}.)\widetilde{X}_1' \\
= \; & (-I_{f,N}+o(1))(\Sigma_2-N^{-1}\mathrm{var}(\widetilde{X}_1)) \\
= \; & -I_{f,N}\Sigma_2+o(1).
\end{aligned}
$$

Note that the first identity follows from (B.9) and the fact that $E\dfrac{f^{(2)}}{f}(W_1)=0$. Since the variance of the second term of (B.19) is of order $1/N$, we established (B.18). ■

B.3 Proof of Theorem 2.3

With all the notations introduced for *Model 1* and in Section 2.2, we can write the log-likelihood of (X,Y) as

$$
\begin{aligned}
L(X, Y; \beta, \rho, \sigma^2, h) &= -\frac{T}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(1-\rho^2) \\
&\quad -\frac{1}{2\sigma^2}\left\{(1-\rho^2)(Y_1 - X_1'\beta)^2 + \sum_{t=2}^{T}(Y_t - X_t'\beta - \rho(Y_{t-1} - X_{t-1}'\beta))^2\right\} \\
&\quad +\frac{W^2}{2v^2} + \log f(W, X) + \frac{1}{2}\log(2\pi v^2),
\end{aligned}
$$

where $f(w, x)$ is the joint density of (W, X) which is given by

$$
\begin{aligned}
f(w, x) &= f(w, x, ; \rho, \beta) \\
&:= \frac{1}{v}\int \phi(\frac{w-u}{v})h(u, x)\, du.
\end{aligned}
$$

The theorem follows by arguments parallel to those for *Model 1* with the facts that (W, X) and $(Z_1, \ldots, Z_T)$ are independent, and that $E\{(f^{(1)}/f)(W, X)|X\} = 0$. ■

References

[1] Adams, R.M., Berger, A.N., and Sickles, R.C. (1997), "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.

[2] Adams, R.M., Berger, A.N., and Sickles, R.C. (1999), "Semiparametric Approaches to Stochastic Panel Frontiers with Applications in the Banking Industry", *Journal of Business and Economic Statistics*, 17, 349–358.

[3] Akhavein, J. D., Swamy, P. V. A. V., and Taubman, S. B. (1997), "A General Method of Deriving the Efficiencies of Banks from a Profit Function", *Journal of Productivity Analysis*.

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

[5] 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.

[6] Berger, A.N. (1993), "Distribution-Free Estimates of Efficiency in U.S. Banking Industry and Tests of the Standard Distributional Assumptions", *Journal of Productivity Analysis*, 4, 261–292.

[7] Berger, A.N., and Humphrey, D.B. (1997), "Efficiency of Financial Institutions: International Survey and Directions for Future Research", *European Journal of Operational Research*, 98, 175–212.

[8] Berger, A.N., Kashyap, A.K., and Scalise J.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.

[9] 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.

[10] Bickel, P.J. and Ritov,Y. (1987), "Efficient Estimation in the Errors in the Variables Model", *Annals of Statistics*, 15, 513–540.

[11] Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moments Restrictions", *Journal of Econometrics*, 34, 305–334.

[12] Chamberlain, G. (1992), "Efficiency Bounds for Semiparametric Regressions", *Econometrica*, 60, 567–596.

[13] 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.

[14] 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.

[15] Fixler, D. and Zieschang, K.E. (1992), "User Costs, Shadow Prices and the Real Output of Banks", in Griliches, Z. (ed.), *Output Measurements in the Service Sectors*, National Bureau of Economic Research Studies in Income and Wealth, Volume 56, Chicago: University of Chicago Press.

[16] Grosskopf, S., Hayes, K.J., Taylor, L.L. and Weber, W.L. (1997), "Budget-Constrained Frontier Measures of Fiscal Equality and Efficiency in Schooling", *Review of Economics and Statistics*, 79, 116–124.

[17] Hausman, J.A. and Taylor, E.E. (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, 49, 1377–1398.

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

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

[20] Humphrey, D.B. and Pulley, L.B. (1997), "Banks' Responses to Deregulation: Profits, Technology, and Efficiency", *Journal of Money, Credit, and Banking*, 73–93.

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

[22] Jayasiriya, R. (2000),*Essays on Structural Modeling Using Nonparametric and Parametric Methods with Applications in the US Banking Industry*, unpublished Ph.D. dissertation, Rice University.

[23] 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.

[24] Lovell, C.A.K., Travers, P., Richardson, S. and Wood, L. (1994), "Resources and Functions: A New View of Inequality in Australia", in Eichhorn, W. (ed.), *Models and Measurements of Welfare and Inequity*, Heidelberg: Physica-Verlag, 643–665.

[25] Newey, W.K. (1990), "Semiparametric Efficiency Bounds", *Journal of Applied Econometrics*, 5, 99–136.

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

[27] Park, B.U. and Simar, L. (1994), "Efficient Semiparametric Estimation in Stochastic Frontier Models", *Journal of the American Statistical Association*, 89, 929–936.

[28] Park, B.U., Sickles, R.C., and Simar, L. (1998), "Stochastic Frontiers: A Semiparametric Approach", *Journal of Econometrics*, 85, 273–301.

[29] Robinson, P. (1988), "Root-N-consistent Semiparametric Regression", *Econometrica*, 56, 931–954.

[30] Schmidt, P. and Sickles, R.C. (1984), "Production Frontiers and Panel Data", *Journal of Business and Economic Statistics*, 2, 367–374.

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

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