"Stochastic panel frontiers: A semiparametric approach"

Park, B. U. ; Sickles, R. C. ; Simar, Léopold

Abstract

This paper complements the results of Hausman and Taylor (1981) and Cornwell, Schmidt and Sickles (1990) and generalizes Park and Simar (1994) by examining the semiparametric efficient estimation of panel models in which the random effects and the regressors have certain patterns of correlation. A model in which this estimator may have particular promise is the stochastic panel frontier model. In that model ineffciency may be correlated with certain determinants of technology or proxies for heterogeneity in the application of that technology. Generalized least squares or other estimators that fail to address this dependency structure are inconsistent. We examine semiparametric effient estimation for three different models based on differing dependency structures. Efficiency of the slope parameters and the asymptotic properties of the level of the frontier function are explored. We illustrate our new estimator in an analysis of productive efficiency between selected North American and...

Document type : *Document de travail (Working Paper)*

Référence bibliographique

Park, B. U. ; Sickles, R. C. ; Simar, Léopold. *Stochastic panel frontiers: A semiparametric approach.* STAT Discussion Papers ; 9601 (1996) 34 pages

STOCHASTIC PANEL FRONTIERS:
A SEMIPARAMETRIC APPROACH

by

B.U. Park*, R.C. Sickles and L. Simar*****

Revised June 1996

Abstract

This paper complements the results of Hausman and Taylor (1981) and Cornwell, Schmidt and Sickles (1990) and generalizes Park and Simar (1994) by examining the semiparametric efficient estimation of panel models in which the random effects and the regressors have certain patterns of correlation. A model in which this estimator may have particular promise is the stochastic panel frontier model. In that model inefficiency may be correlated with certain determinants of technology or proxies for heterogeneity in the application of that technology. Generalized least squares or other estimators that fail to address this dependency structure are inconsistent. We examine semiparametric efficient estimation for three different models based on differing dependency structures. Efficiency of the slope parameters and the asymptotic properties of the level of the frontier function are explored. We illustrate our new estimator in an analysis of productive efficiency between selected North American and European airline firms after domestic deregulation in the U.S. and prior to recent European reforms implemented in the course of EC integration.

Key words: Stochastic frontier model, semiparametric model, efficient estimation, information bound, panel data.

[0]* Department of Computer Sciences and Statistics, Seoul National University, Korea.
**Department of Economics, Rice University, U.S.A.
*** CORE and Institute of Statistics, Université Catholique de Louvain, Louvain-la-Neuve.
Research support from the contract "Projet d'Actions de Recherche ConcertÇes" and the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State are gratefully acknowledged.

1 Introduction[1]

In this paper, we analyze the semiparametric efficient estimation of stochastic panel frontiers in which the random effects (capturing economic inefficiency of the firms) and the regressors have certain dependency structures. We extend the Park and Simar (1994) model wherein the effects and regressors were assumed to be independent by considering three classes of error and regressor dependency. The first assumes no particular structure of dependence between the effects and the regressors. In the second, structure is imposed by allowing dependency between the effects and a subset of regressors. The third allows for dependency between the effects and long run movements in a subset of regressors. Although the models we consider are intended to apply to generic panel data problems, an appealing empirical motivation for the appropriate treatment of them is in the estimation of firm specific efficiency levels in a stochastic panel production frontier model (c.f. Schmidt and Sickles, 1984; Cornwell, Schmidt and Sickles, 1990).

Section 2 discusses the three alternative models considered in our paper and derives the asymptotic lower bound and the efficient estimator for different dependency structures, termed **Models 1-3**. It appears that the within estimator is efficient for the general setup of **Model 1.** However, due to the structure of dependence in **Model 2** and **Model 3**, the within estimator can be improved upon. Under the setup of **Model 2**, which is our main contribution, we propose in section 3 predictors of the individual effects whose construction utilizes the efficient estimator of the slope parameters θ and provide proofs of their asymptotic properties. We also discuss an estimator of the level of the frontier function B, which is the upper boundary of the support of the marginal density of the effects α, and examine its rate of convergence to the desired quantity. It is this boundary estimator on which is based the estimates of firm specific relative technical efficiency. We are unaware of any asymptotic results (beyond consistency) involving the level of the frontier function when such dependence structure between regressors and effects exist. Section 4 uses the **Model 2** estimator to illustrate our semiparametric paradigm by studying efficiency differentials between North American and European airline firms during the period 1976-1990. Section 5 concludes. In appendix 1 we present briefly the general theory of efficient estimation of slope parameters in the presence of nuisance parameters. Proofs of major theorems are in appendix 2.

2 Efficiency Bound and Efficient Estimators

The basic model we analyze, under its more general form, can be written as:

$$Y_{it} = X_{it}'\theta + \alpha_i + \varepsilon_{it} \qquad i = 1, \ldots, N; t = 1, \ldots, T \tag{2.1}$$

[1]The authors would like to thank Robert Adams for his valuable contributions to this and earlier versions of the paper.

where $X_{it} \in \mathbb{R}^d$, $\theta \in \mathbb{R}^d$. The ε_{it} are assumed to be iid random variables from a $N(0, \sigma^2)$. Let $X_i = (X'_{i1}, \ldots, X'_{iT})'$. We assume that the (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 assumed to 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. For the stochastic panel production frontier model 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 return to this issue below.[2]We consider below three different structures of dependency between X and α. **Model 1** refers to equation (1.1) when no particular structure of dependence is assumed between X and α. **Model 2** refers to the case in which there is dependency between a subset of regressors (Z) and the effects. The final model we analyze is one in which the dependence between Z and α is through long run movements in Z, that is, Z depends on α through $\overline{Z}$. This is our **Model 3.**

The choice of which variables should be included in Z_{it} (**Model 1** vs. **Model 2/3**) and the type of dependency structure for (α, z) (**Model 2** vs. **Model 3**) is contingent on the model and empirical setting. In the context of the stochastic panel frontier production function, specification of **Model 1** would lead to a semi-parametric efficient estimator analogous to the dummy variable or fixed effects estimator where no orthogonality restrictions are imposed between regressors and effects and thus no time invariant regressors can be identified. **Models 2 and 3** lead to semiparametric efficient estimators which are analogous to the estimator of Hausman and Taylor (1981). **Model 2** allows for general dependencies between the effects and the regressors with which they are correlated, while **Model 3** allows dependencies only between the effects and average (long run) movements in the regressors with which they are correlated. In a stochastic frontier production function setting, this last dependency structure may be appropriate if the systematic shortfall in production (identified by the firm effect) due to misapplication of technology by a firm, or to the presence of mismeasured binding constraints which interfere with the application of best practice technology, tends to rise or fall only when accompanied by systematic (long run) changes in factors which contribute

[2]A referee has pointed out two promising alternatives to our model in which the parametric assumption of the normality of the population disturbance, the "within" error, is replaced by either a pure independence assumption yielding an adaptive estimator which is efficient or a zero conditional mean assumption yielding a weighted least squares version of the within component. Utilizing the latter assumption also would require nonparametric estimation of the conditional variance (Robinson, 1988).

to its presence in the first place.

Throughout this section except in the discussion concerning Theorem 3.4, the time period T is considered fixed. We now proceed to derive the asymptotic lower bound for estimating $\theta = (\beta, \gamma)'$ in the presence of nuisance parameters σ^2 and h for our three classes of dependency structure. The efficient semiparametric estimator of θ attains this lower bound.

2.1 Model 1

Now we will make explicit the information bound for the **Model 1**. Recall that **Model 1** is:

$$Y_{it} = X_{it}'\theta + \alpha_i + \varepsilon_{it} \qquad i = 1, \dots, N; t = 1, \dots, T \tag{2.2}$$

Let $Y = (Y_1, \dots, Y_T)'$, $X = (X_1', \dots, X_T')$ for the generic observation (X, Y). Let $S_t(\theta) = Y_t - X_t'\theta$ and $U_t(\theta) = S_t(\theta) - \overline{S}(\theta)$ where $\overline{S}(\theta) = T^{-1}\sum_{t=1}^{T} S_t(\theta)$. Similarly, $\overline{X} = T^{-1}\sum_{t=1}^{T} X_t$, $\overline{Y} = T^{-1}\sum_{t=1}^{T} Y_t$ and $\overline{\sigma} = \sigma/\sqrt{T}$. Let $\Sigma_W = E(T^{-1}\sum_{t=1}^{T}(X_t - \overline{X})(X_t - \overline{X})')$ which is assumed to exist and to be nonsingular. Utilizing the basic results on construction of semiparametric efficient estimation discussed in appendix 1 as well as the notation contained therein, we have the following theorem:
Theorem 2.1

The efficient score function and the information bound for estimating θ in **Model 1** are given by

$$\ell^* = \sigma^{-2}\sum_{t=1}^{T} U_t(\theta) X_t$$
$$I(P; \theta, \mathbb{P}) = \overline{\sigma}^{\,-2}\Sigma_W$$

The proof is in the appendix. The efficient estimator of θ for **Model 1** is given by the following:

Corollary 2.1

The *within estimator*

$$\tilde{\theta}_N = (NT\hat{\Sigma}_W)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \overline{X}_i)(Y_{it} - \overline{Y}_i)$$

where $\hat{\Sigma}_W = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \overline{X}_i)(X_{it} - \overline{X}_i)'$

is efficient in **Model 1**.

Proof: The *within estimator* is the OLS estimator of θ obtained by regressing $(Y_{it} - \overline{Y})$ on $(X_{it} - \overline{X})$. It is indeed straightforward to show that as $N \to \infty$, we have

$$\mathcal{L}_P(\sqrt{N}(\tilde{\theta}_N - \theta)) \to N(0, \overline{\sigma}^{\,2}\Sigma_W^{-1})$$

Remark 2.1. When there is no correlation between the effects and the regressors the within estimator is no longer semiparametric efficient. In this case the semiparametric efficient estimator for the pure random effects model is the one derived in Park and Simar (1994).

2.2 Model 2

This model can be written as:

$$Y_{it} = X_{it}'\beta + Z_{it}'\gamma + \alpha_i + \varepsilon_{it} \tag{2.3}$$

where $X_{it} \in \mathbb{R}^p$ and $Z_{it} \in \mathbb{R}^q$ where $d = p + q$. Given $Z_i = (Z_{i1}', \ldots, Z_{iT}')'$, we add a further assumption that α_i and X_i are conditionally independent. In other words, the density of the (α_i, Z_i, X_i)'s may be written

$$f(\alpha, z, x) = h(\alpha, z)\; g(x|z) \tag{2.4}$$

where $h(\cdot, \cdot)$ is the unknown joint density of (α_i, Z_i) and $g(\cdot|\cdot)$ is the conditional density of X_i given Z_i. As in **Model 1**, the support of the marginal density of α is bounded above (or below). We see from (2.2) that Z and α may be dependent but the dependence between X and α passes through Z.

Consider as above the generic observation (X, Z, Y). Here $S_t(\theta) = Y_t - X_t'\beta - Z_t'\gamma$, again $U_t(\theta) = S_t(\theta) - \overline{S}(\theta)$ and $\overline{Z} = T^{-1}\sum_{t=1}^{T} Z_t$. Let

$$w(s, z) = \int \phi_{\overline{\sigma}}(s - u)h(u, z)du = \phi_{\overline{\sigma}} * h(\cdot, z)(s)$$

be the joint density of $(\overline{S}(\theta), Z)$ on $\mathbb{R}^{1+Tq}$ and

$$w'(s, z) = \frac{\partial}{\partial s}w(s, z).$$

Let

$$I_0 = \int \frac{(w')^2}{w}(s, z)dsdz$$

Define

$$\Sigma_W(X) = E(T^{-1}\sum_{t=1}^{T}(X_t - \overline{X})(X_t - \overline{X})')$$
$$\Sigma_W(Z) = E(T^{-1}\sum_{t=1}^{T}(Z_t - \overline{Z})(Z_t - \overline{Z})')$$
$$\Sigma_W(X, Z) = E(T^{-1}\sum_{t=1}^{T}(X_t - \overline{X})(Z_t - \overline{Z})')$$
$$\Sigma_B(X|Z) = E((\overline{X} - E(\overline{X}|Z))(\overline{X} - E(\overline{X}|Z))')$$

Throughout this paper we will assume that $I_0 < \infty$ and that the within moment matrices Σ_W and the between moment matrix exist and are nonsingular.

Theorem 2.2

Let $\ell^* = (\ell_\beta^{*'}, \ell_\gamma^{*'})'$. The efficient score function and the information bound for **Model 2** are given by

$$\ell_\beta^* = \sigma^{-2}\sum_{t=1}^{T} U_t(\theta)X_t - (\overline{X} - E(\overline{X}|Z))\frac{w'}{w}(\overline{S}(\theta), Z)$$
$$\ell_\gamma^* = \sigma^{-2}\sum_{t=1}^{T} U_t(\theta)Z_t$$

$$I(P;\theta,\mathbb{P}) = \begin{pmatrix} \overline{\sigma}^{-2}\Sigma_W(X) + I_0\Sigma_B(X|Z) & \overline{\sigma}^{-2}\Sigma_W(X,Z) \\ \overline{\sigma}^{-2}\Sigma_W'(X,Z) & \overline{\sigma}^{-2}\Sigma_W(Z) \end{pmatrix}$$

The proof is in the appendix.

Remark 2.2. The normality assumption simplifies the derivation of the semiparametric efficient estimator and allows us to focus our attention on two questionable assumptions often made in the production efficiency literature, specific *a priori* parametric forms for the distribution of inefficiency and independence of ineffiency and regressors. Moreover, we are unaware of any stochastic frontier models which utilize parametric distributions for the disturbance other than the normal (see, e.g., the extensive surveys by Schmidt, 1985; Bauer, 1990; Greene, 1996). However, as pointed out in section 1, this assumption could in principle also be lifted. To outline one approach for such an estimator consider the model of Park and Simar (1994). In addition to the independence assumption between x_{it}, ϵ_{it}, and α_i, let us assume that $(\epsilon_{it} - \bar{\epsilon}_i, ..., \epsilon_{iT} - \bar{\epsilon}_i)$ is independent of $\alpha_i + \bar{\epsilon}_i$. Let g, f_w, and f_b be the density of $(X_1, ...X_T)$, $(\epsilon_{it} - \bar{\epsilon}_i, ..., \epsilon_{iT} - \bar{\epsilon}_i)$, and $\alpha_i + \bar{\epsilon}_i$, respectively.

Then the log likelihood of the generic observation $(X,Y) \equiv (X_1,...,X_T,Y)$ is:

$$L(Y,X;\beta,f_w,f_b) = \log\ g(X)\ +\ \log\ f_w(U_1(\beta),...,U_T(\beta)) + \log\ f_b(\bar{S}(\beta))$$

The efficient score for estimating β can be shown to be

$$\ell^*_\beta\ =\ -\left\{X - E(X) - (\overline{X} - E(\overline{X}))\right\}^T \frac{f'_w}{f_w}(U_1(\beta),...,U_T(\beta))\ - (\bar{X} - \mathrm{E}(\bar{X}))\frac{f'_b}{f_b}(\bar{S}(\beta))$$

where $\{\cdot\}^T$ denotes the transpose. The information bound then can be obtained from this efficient score.[3]

Remark 2.3. When h and σ^2 are known, the $\Sigma_B(X|Z)$ in the information bound is replaced by $E\overline{X}\ \overline{X}'$.

Remark 2.4. It can be seen that the information bound I in Theorem 2.2 coincides with $I(P;\theta,\mathbb{P}_0)$ where $\mathbb{P}_0$ is a regular parametric submodel obtained by parameterizing the density of (α,Z) by $h_{\mu,\eta}(\alpha,z) = h(\alpha-\mu-\eta' z, z)$ where $h(\alpha,z)$ is the fixed density of (α,Z) under P. This is the *least favorable* regular parametric submodel of $\mathbb{P}$.

For later use let us define

$$\Sigma_W = \begin{pmatrix} \Sigma_W(X) & \Sigma_W(X,Z) \\ \Sigma'_W(X,Z) & \Sigma_W(Z) \end{pmatrix}$$

We now construct an efficient estimator of θ. The idea is to proceed as in the classical estimation of the location problem: Let $\tilde{\ell} = I^{-1}\ell^*$. This is called the *efficient influence function* having the properties $E[\tilde{\ell}] = 0$ and $E[\tilde{\ell}\tilde{\ell}'] = I^{-1}$.

(a) Find a consistent estimator $\tilde{\theta}_N$ of θ.

(b) Consider $\tilde{\ell}$ as $\tilde{\ell}(x,z,y,\theta,\sigma^2,h,g)$. Construct a suitable estimator $\tilde{\ell}(\cdot,\cdot,\cdot,\theta;X_1,Z_1,Y_1,\dots,X_N,Z_N,Y_N)$ of $\tilde{\ell}(\cdot,\cdot,\cdot,\theta,\sigma^2,h,g)$.

(c) Form

$$\hat{\theta}_N = \tilde{\theta}_N + N^{-1}\sum_{i=1}^{N}\tilde{\ell}(X_i,Z_i,Y_i,\tilde{\theta}_N;X_1,Z_1,Y_1,\dots,X_N,Z_N,Y_N)$$

[3] We test the within residuals for departures from normality in the empirical illustration in section 5 below and find no evidence of nonnormality.

as the efficient estimator. The above construction is motivated by the fact that

$$N^{-1/2}\sum_{i=1}^{N}\tilde{\ell}(X_i, Z_i, Y_i, \theta, \sigma^2, h, g)$$

is asymptotically $N(0, I^{-1})$.

For a preliminary estimator $\tilde{\theta}_N$ of θ, we can use the within estimator obtained from regressing $Y_{it} - \overline{Y}_i$ on $X_{it} - \overline{X}_i$ and $Z_{it} - \overline{Z}_i$ by the ordinary least squares method. In fact,

$$\tilde{\theta}_N = \begin{pmatrix} \tilde{\beta}_N \\ \tilde{\gamma}_N \end{pmatrix} = (NT\hat{\Sigma}_W)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{pmatrix} X_{it} - \overline{X}_i \\ Z_{it} - \overline{Z}_i \end{pmatrix}(Y_{it} - \overline{Y}_i)$$

where $\hat{\Sigma}_W$ is the sample version of Σ_W. As pointed above, as $N \to \infty$ (with fixed T), we have

$$\mathcal{L}_P(\sqrt{N}(\tilde{\theta}_N - \theta)) \to N(0, \overline{\sigma}^2\Sigma_W^{-1}).$$

In view of Theorem 2.2, the within estimator is not efficient. However, it already lies in a $N^{1/2}$-neighborhood of θ, which is enough for the final estimator given below to be efficient. An alternative preliminary estimator of θ is the Hausman and Taylor (1981) efficient instrumental variables estimator[4]. Let M project the data into individual means and Q project it into deviations from individual means, i.e. $M = I_N \otimes e_T e_T^{'}/T$, $Q = I_{NT} - M$, where e_T is a $(T \times 1)$ column vector of ones. If we use the standard stacking conventions for the data in Y_{it}, X_{it}, Z_{it} so that the first T observations contain the time series for the first cross-sectional unit and so on, then the within estimator for θ is $\hat{\theta} = [(X, Z)'Q(X, Z)]^{-1}\ (X, Z)'QY$. The Hausman and Taylor estimator is

$$\hat{\theta} = [(X^*, Z^*)'(X^*, \hat{Z}^*)]^{-1}(X^*, \hat{Z}^*)'Y^*$$

where, e.g. $Y_{it}^* = [Y_{it} - (1 - \sigma_W/\sqrt{T}\sigma_B)\overline{Y}_i]$ and the instrument set for $\hat{Z}$ is $[QX, QZ, MX]$. In view of the efficiency gain due to the use of additional instruments MX, the Hausman-Taylor estimator is more efficient than within, but in light of Theorem (2.2) it is not efficient. It can serve, however, as an alternative root-N consistent estimator on which the efficient estimator is based[5].

To construct the efficient estimator, define an estimator of w by the $(1 + Tq)$ dimensional kernel estimator,

$$\hat{w}(s, z, \theta) = N^{-1}\sum_{i=1}^{N}K_{s_N}(s - \overline{S}_i(\theta))\prod_{t=1}^{T}\prod_{k=1}^{q}K_{s_N}(z_{tk} - Z_{itk}) + c_N$$

[4]Other consistent first step estimators could also be used, e.g. Amemiya and McCurdy (1986) and Breusch, Mizon, and Schmidt (1989).

[5]Of course the within estimator could be preferred on the gounds of robustness with respect to misspecified orthogonality between effects and selected regressors and/or low power of standard tests to detect such misspecified orthogality.

where $K_s(t) = K(t/s)/s,\ K(t) = e^{-t}(1+e^{-t})^{-2}$, and s_N, c_N tend to zero at some proper rates as specified below. Note that we use the same kernel function K and the bandwidth s_N for both $\overline{S}$ and Z_{tk} $(t = 1, \ldots, T; k = 1, \ldots, q)$. One may use two different kernels and bandwidths here, but we avoid this just for the simplicity of presentation. A straightforward extension to the latter case is possible.

Define

$$\hat{\sigma}^2(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}^2(\theta)/N(T-1)$$

To estimate at $\sqrt{N}$-rate $E(\overline{X}|Z)$ in ℓ_β^* and $\Sigma_B(X|Z)$, we need a parametric form for $E(\overline{X}|Z)$. So, we assume, for example[6],

$$E(\overline{X}|Z) = a + AZ$$

where a is a $p \times 1$ vector and A is a $p \times Tq$ matrix.

Then, one can estimate a and A at the $\sqrt{N}$-rate by the standard least squares methods. Call the estimates by $\hat{a}$ and $\hat{A}$ and let $p(Z) = a + AZ$, $\hat{p}(Z) = \hat{a} + \hat{A}Z$. Define

$$\hat{\Sigma}_B(X|Z) = \frac{1}{N} \sum_{i=1}^{N} (\overline{X}_i - \hat{p}(Z_i))(\overline{X} - \hat{p}(Z_i))'$$

The efficient estimator, as prescribed above is now defined by

$$\hat{\theta}_N = \tilde{\theta}_N + N^{-1}\hat{I}^{-1} \begin{bmatrix} \sum_{i=1}^{N} \left\{ \sum_{t=1}^{T} \tilde{U}_{it} X_{it}/\tilde{\sigma}^2 - (\overline{X}_i - \hat{p}(Z_i)) \frac{\hat{w}'}{\hat{w}} (\tilde{S}_i, Z_i, \tilde{\theta}_N) \right\} \\ \\ \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{U}_{it} Z_{it}/\tilde{\sigma}^2 \end{bmatrix}$$

where

$$\hat{I} = \begin{bmatrix} T\hat{\Sigma}_W(X)/\tilde{\sigma}^2 + \hat{I}_0 \hat{\Sigma}_B(X|Z) & T\hat{\Sigma}_W(X,Z)/\tilde{\sigma}^2 \\ \\ T\hat{\Sigma}_W(X,Z)'/\tilde{\sigma}^2 & T\hat{\Sigma}_W(Z)/\tilde{\sigma}^2 \end{bmatrix}$$

$$\hat{I}_0 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{w}'}{\hat{w}} \right)^2 (\tilde{S}_i, Z_i, \tilde{\theta}_N)$$

and $\tilde{\sigma}^2$, $\tilde{U}_{it}$, $\tilde{S}_i$ are used for $\hat{\sigma}^2(\tilde{\theta}_N)$, $U_{it}(\tilde{\theta}_N)$, $\overline{S}_i(\tilde{\theta}_N)$.

Let $\{c_N\}$, $\{s_N\}$ be such that

$$c_N \to 0, \quad s_N \to 0, \quad N c_N^2 s_N^6 \to \infty$$

[6]Higher order polynomial in Z or other parametric forms could also be used but for simplicity we propose a linear model for $E(\overline{X}|Z)$.

as $N \to \infty$. Let h_1 and h_2 denote the marginal density of α and Z respectively.

Theorem 2.3. Assume $E(e^{t|\overline{X}|}) < \infty$ for some $t > 0$, and $\int |u|^2 h_1(u)du < \infty$. If h_1, h_2 are bounded, then

$$\mathcal{L}_P(\sqrt{N}(\hat{\theta}_N - \theta)) \to N(0, I^{-1}).$$

The speed of the bandwidth s_N required for Theorem 2.3 is not dependent on the value of T and q since how fast s_N should go to zero is determined by the value k such that

$$\sup_{s,z,\theta} s_N^k \left| \frac{\hat{w}^{(j)}}{\hat{w}}(s, z, \theta) \right| < \infty \tag{2.5}$$

where, $\hat{w}^{(j)}(s, z, \theta) = \dfrac{\partial^j}{\partial s^j} \hat{w}(s, z, \theta)$ and it is easy to see that the value of k ensuring (3.5) is $k = j$.

Remark 2.5. The conditions imposed on h in the statement of the above theorem do not affect the form of the information bound in Theorem 2.1 if the underlying probability distribution P belongs to the reduced model $\mathbb{P}_1(\subset \mathbb{P})$ induced by those conditions and on extra one $\int | z |^2 h_2(z)dz < \infty$. This is clearly seen from the fact that the least favorable regular parametric submodel $\mathbb{P}_0$ described in Remark 2.3 is contained in $\mathbb{P}_1$ so that $I(P; \theta, \mathbb{P}_1) = I(P; \theta, \mathbb{P})$.

It would be interesting to investigate the asymptotic properties of $\hat{\theta}_N$ when T also goes to infinity. For this, we need to assume that $R_{it} = (X'_{it}, Z'_{it})'$ are i.i.d. $p + q$-dimensional random vectors from unknown density function g_1. Append the subscript T (in addition to N) in all the corresponding sequences of random vectors and numbers. The following theorem will be useful in the next section. It requires additional assumptions on the consistency of the first step estimator $\tilde{\theta}_{N,T}$ which are shared by the *within* estimator of θ.

Theorem 2.4. Let R denote the generic observations R_{it}. Suppose $E|R|^2 < \infty$ and $E(R - E(R))(R - E(R))'$ is nonsingular.
(i) If

$$\sqrt{T}(\tilde{\theta}_{N,T} - \theta) = O_p(1) \text{ when } N \text{ is fixed and } T \to \infty \tag{2.6}$$

then, when N fixed, $T \to \infty$ and $Ts_{N,T}^2 \to \infty$,

$$\sqrt{T}(\hat{\theta}_{N,T} - \theta) = O_p(1)$$

(ii) Further if

$$\sqrt{(NT)}(\tilde{\theta}_{N,T} - \theta) = O_p(1) \text{ when both } N, T \to \infty \tag{2.7}$$

then, when $N, T \to \infty$, $Ts_{N,T}^2 \to \infty$ and $NT^{-1}s_{N,T}^{-2} = O(1)$,

$$\sqrt{NT}(\hat{\theta}_{N,T} - \theta) = O_p(1).$$

2.3 Model 3

This model differs from **Model 2** by the addition of the following restriction on the joint density of (α, z) :

$$h(\alpha, z) = h_M(\alpha, \overline{z})p(z) \tag{2.8}$$

The theory is essentially the same as in **Model 2** but due to the additional restriction (2.3) , we have here to consider

$$w(s, \overline{z}) = \int \phi_{\overline{\sigma}}(s-u)h_M(u, \overline{z})du,$$

the density of $(\overline{S}(\theta), \overline{Z})$. As in section 2.2. we have the theorem analog to theorem 2.2. Let

$$\Sigma_B(X|\overline{Z}) = E(T^{-1}(\overline{X} - E(\overline{X}|\overline{Z}))(\overline{X} - E(\overline{X}|Z))')$$

and as above let

$$I_0 = \int \frac{[w'(s, \overline{z})]^2}{w(s, \overline{z})} ds d\overline{z}$$

We have the following result:

Theorem 2.5. The efficient score function, $\ell^* = (\ell_\beta^{*'}, \ell_\gamma^*)$, and the information bound for **Model 3** are given by:

$$\ell_\beta^* = \sigma^{-2}\sum_{t=1}^{T} U_t(\theta)X_t - (\overline{X} - E(\overline{X}|\overline{Z}))\frac{w'}{w}(\overline{S}(\theta), \overline{Z})$$
$$\ell_\gamma^* = \sigma^{-2}\sum_{t=1}^{T} U_t(\theta)Z_t$$

$$I(P;\theta,\mathbb{P}) = \begin{pmatrix} \overline{\sigma}^{-2}\Sigma_W(X) + I_0\Sigma_B(X|\overline{Z}) & \overline{\sigma}^{-2}\Sigma_W(X,Z) \\ \overline{\sigma}^{-2}\Sigma'_W(X,Z) & \overline{\sigma}^{-2}\Sigma_W(Z) \end{pmatrix}$$

The efficient estimator in this case may be derived, *mutatis mutandis*, as in **Model 2**.

3 Individual effects and level of frontier function

In this section, we analyze, under the setup of **Model 2**, the predictors of the individual effects and the estimator of the level of frontier function. The adaptation for the other models is straightforward. We append the subscript T in all the related quantities to stress their dependence on T. Given the efficient estimator $\hat{\theta}_{N,T}$ at hand, it would be natural to predict α_i by

$$\hat{\alpha}_i = \overline{S}_i(\hat{\theta}_{N,T}).$$

In this section, we investigate how fast $\hat{\alpha}_i - \alpha_i$ goes to zero as T (and/or N) goes to infinity. Note that, with fixed T, it is easy to see, from the results of Section 3,

$$\hat{\alpha}_i - \alpha_i = O_p(N^{-1/2})$$

for each i. As in Theorem 2.4, we need to assume that R_{it} are i.i.d. $p+q$-dimensional vectors from unknown density g_1. In the following theorem, N may be fixed or tend to infinity.

Theorem 3.1. Under the assumptions of Theorem 2.4., as $T \to \infty$ and $Ts_{N,T}^2 \to \infty$,

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

Remark 3.1. When one is interested in predicting the difference $\alpha_i - \alpha_j$ (the relative technical inefficiencies of the i-th firm with respect to the j-th firm), one can predict this by $\hat{\alpha}_i - \hat{\alpha}_j$. This predictor has the asymptotic $N(0, 2\sigma^2)$ distribution when normalized by $\sqrt{T}$ under the same assumptions of the above theorem.

If we let B denote the upper boundary of the support of h_1, then the frontier function will be $B + x'\beta + z'\gamma$. Since we already have an efficient estimator of θ, it remains to estimate B to get an estimator of this function. A natural estimator of this quantity would be

$$\hat{B} = \max_{1 \le i \le N} \overline{S}_i(\hat{\theta}_{N,T}).$$

The following two theorems will be useful to investigate the asymptotic property of this estimator. For this let $\alpha_{(N)} = \max\limits_{1 \le i \le N} \alpha_i$.

Theorem 3.2. Under the assumptions of Theorem 2.4 and if $E(e^{t|R|}) < \infty$ for some $t > 0$, as $T \to \infty$ and $Ts_{N,T}^2 \to \infty$,

(i) when N is fixed,

$$\sqrt{T}(\hat{B} - \alpha_{(N)}) = O_p(1),$$

(ii) when N goes to infinity,

$$\sqrt{T}(\hat{B} - \alpha_{(N)}) = O_p(\log N).$$

How close $\alpha_{(N)}$ is to B is illustrated by the following theorem. For this, suppose the density h_1 satisfies

$$h_1(B - u) \ge Cu^\delta \tag{3.9}$$

as $u \downarrow 0$ for some constant $C > 0$ and $\delta \ge 0$. This condition requires that the density h_1 should have a certain amount of mass near the boundary point B.

Theorem 3.3. Under the assumptions of Theorem 2.4 and the condition (3.9), as $N \to \infty$,

$$\alpha_{(N)} - B = O_p(N^{-\frac{1}{\delta+1}}).$$

The implication of Theorem 3.2 and 3.3 is clear. For example, when $\delta = 0$ (the case where the density at the boundary stays away from zero, such as shifted half-normal and exponential),

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

if both N and T go to infinity.

4 An empirical illustration

We next implement the semiparametric efficient estimators in the stochastic panel frontier model using data from fifteen North American and European airline firms during the period 1976-1990 to ascertain the relative efficiency of particular carriers as well as the relative efficiency of the North American and European airline industries. One might expect that inefficiencies existed in the European airline industry during this period while the regulatory reforms from European Union were not yet in place but those from U. S. airline deregulation largely were. We specify a Cobb-Douglas production function in which the natural logarithm of output is specified to be a linear function of the natural logarithms of factor inputs. As with Schmidt and Sickles (1984) and Cornwell et al. (1990) we interpret the firm effect in the panel production function as technical inefficiency.

The airline data set consists of an annual panel of the eight largest European carriers (Air France, Alitalia, British Airways, Iberia, KLM, Lufthansa, SAS, and Sabena) and seven American carriers (American, Continental, Delta, Eastern, Northwest, TWA, and United) which operated during the period 1976 and 1990 and which supply virtually all of the scheduled international traffic of the U.S. airline industry. The primary source for our data is the *The Digest of Statistics* from the International Civil Aviation Organization (ICAO). Using this source, we construct a set of four airlines inputs: Labor, Energy, Materials and Aircraft Fleet. In addition, aggregate airline outputs and several of its characteristics are constructed. Details of data construction for a subset of carriers and a shorter panel can be found in Good et al. (1993).

The labor input is an aggregate of five separate categories of employment used in the production of air travel. These categories include pilots (as well as copilots and other cockpit crew), flight attendants, mechanics, ticketing and passenger handlers, and other employees. A Tornqvist multilateral index number is used to aggregate these subcomponents.

Expenditures on supplies, services, ground based capital equipment and energy are combined into a single residual input aggregate called materials. Purchasing power parity exchange rates are adjusted by allowing for changes in market exchange rates and changes in price levels. Fuel expenses are given for each carrier in ICAO's *Financial Data Series*. The airline's fuel price is estimated as a weighted average of the domestic fuel price (weighted by domestic available ton-kilometers), and regional prices (weighted by international available ton-kilometers in the relevant region). These subcomponents are aggregated using the same multilateral index number procedure and are termed materials.

Detailed information for aircraft fleets is provided in ICAO's *Fleet and Personnel Series*. In addition to counts of the total number of aircraft, we construct the percentage of the fleet which is widebodied (a measure of average equipment size) and the percentage of the fleet which is turboprop (a measure of aircraft speed).

Information regarding the carrier's physical outputs are obtained from ICAO's *Commercial Airline Traffic Series*. We consider three subcomponents of airline output: passenger service, cargo operations, and incidental services (which includes equipment leasing, maintenance provided to other carrier's equipment, etc.). Revenues for passenger, cargo and incidental outputs are obtained in ICAO's *Financial Data Series*. The country's purchasing power parity is used as a price deflator for incidental output. Revenue ton-kilometers is used as a quantity deflator for passenger and cargo outputs. These three components are aggregated into a single output using a multilateral index number procedure. Capacity ton miles is then derived as the ratio of revenue ton mile and passenger load factor. Two characteristics of airline output are also calculated - average stage length which provides a measure of the length of individual route segments in the carrier's network and the number of route kilometers which provides a measure of total network size.

The specific form for the Cobb-Douglas production function has $Y = \ln$ (capacity ton mile service); $X = \{\ln$ (capital), ln (materials), percentage of the fleet turboprop, percentage of the fleet wide bodied, ln (average stage length), ln (network size), ln (loadfactor), time$\}$. For **Model 1** [eq. (1.1)], whose semiparametric efficient estimator is *within,* we include ln(labor) in the X vector. For **Models 2 and 3** [eq. (1.2)] we focus on correlation between the inefficiency effects and the single regressor $Z = \ln(\text{labor})$. Roeller and Sickles (1994) examined patterns of misallocation of factor inputs for the European airline industry and found that there was substantial over-use of labor by the European carriers due to the presence of strong unions and the difficulty with which those firms could lay-off workers, a fact brought home by the recent (1993) aborted attempt by Air France management to reduce its labor force. We also focus on a single regressor due to the "curse of dimensionality" problem of multivariate kernel density estimation in higher dimensions than those examined herein. Product kernels can be specified in principle for arbitrary numbers of correlated regressors. Estimation of two-dimensional kernels is possible with our cross-section of 15 firms, but estimation of higher dimensional kernels does not appear to be. The dimensionality problem is also a constraining factor in the implementation of the semiparametric efficient estimator for **Model 2** since it requires a (1+Tq) dimensional kernel estimator, which for the international airline example is a sixteen dimensional kernel estimator (T=15, q=1). **Model 3**, however, in which the effects α (technical efficiencies) are correlated with $\overline{Z}$ (long run changes in ln labor), can be implemented since it only requires a (1+q)=2 dimensional kernel estimator. Clearly, the types of data sets which would be most promising for the **Model 2** estimator in frontier productive efficiency modeling would be ones in which the number of firms was large relative to the number of time periods. Public use data of this sort is not widely available. The applicability of the large sample theory developed in the earlier sections should thus be viewed with some skepticism and thus our results should be viewed as illustrative.[7]

Efficient semiparametric estimates and standard errors for **Model 1** (*within*) and for **Model 3** (along with the first step Hausman-Taylor estimates) for the stochastic panel frontier production function are provided in Table 1.[8] [9] A plot of the within residuals in Figure **??** reveals a strong normal pattern which is confirmed by a Kolmogorov-Smirnov (KS) statistic for the within residuals of 0.0409 ($KS_{0.05} = 0.0907$). We base the efficient estimator of θ on a bandwidth s chosen by a bootstrap procedure. The bootstrap choice of s is $s^* = \arg_s \min C(s)$ where

$$C(s) = (1/M) \sum_{m=1}^{M} \left[\hat{\theta}^{*(m)}_{N,T}(s) - \hat{\theta}_{N,T}(s)\right]' \left[\hat{\theta}^{*(m)}_{N,T}(s) - \hat{\theta}_{N,T}(s)\right]$$

and $\hat{\theta}^{*(m)}_{N,T}(s)$ denotes the m-th pseudo sample bootstrap version of $\hat{\theta}_{N,T}(s)$ using bandwidth s. Our estimates are based on $M = 1000$ and consisted of a grid search in the interval $[0.1, 2.0]$. The optimal value of s^* was 0.271[10]. We then reestimated θ using this selected bandwidth on the original data.

As Table 1 indicates, output elasticities of the factor inputs are stable across the different estimators. Technology was assumed to have long run constant returns to scale with respect to these factor inputs. The test of constant returns to

[7] A referee has pointed out that comparability of the number of regression coefficients to both N and T may explain why the additional "between" component of the efficient estimator doesn't add much to the results. We have conducted a limited Monte Carlo experiment based on **Model 3** using regressors and correlation structures similar to those in our empirical illustration. In these experiments we vary N and T from 15 to 500 and utilize both our kernel estimator as well as higher order bias reducing kernels. There is no evidence of bias with sample sizes of N=15 and T=15. Root mean squared error and absolute mean deviations based on 500 simulations range from 10% to 1% of the value of the parameter estimates. These fall to between 7% to .6% for N=50 and T=15. Bias-reducing kernels did not appear to have any systematic advantage over other kernels.

[8] As pointed out by referees and as discussed in Sections 3 and 4, our semi-parametric efficient estimator requires large N and T for both consistency of the boundary function which allows us to identify the most efficient firm and of our coefficient estimates. Our estimates should thus be viewed as illustrative. However they also point to inefficiency gaps between Europe and American firms which are in keeping with the 10%-15 % cost reductions recently put in place (or attempted) by a number of the European firms in our sample. Our estimator also requires independence of different cross-section observations. This may be particularly suspect in the European industry were collusion may vitiate the integrity of this assumption. However, in recent work by Roeller and Sickles (1994) using a two stage product-differentiated game, the European industry was found to have relatively little market power due to collusion. In fact, firms compete significantly more than a standard Nash game would suggest. On the other hand, significant monopoly power is identified in domestic markets. Price mark-ups appear to be due to inefficiency which firms can pass on to consumers as well as to market niches.

[9] A referee has raised the question of why firm effects and not time effects are emphasized in our empirical illustration. Our motivation is basically an empirical one. Earlier work by Sickles (1985) and Good et al. (1993) have found time effects to be of marginal significance and neglible relative to firm effects for both the U.S.and European carriers.

[10] The average standard deviation for $\overline{S}$ and $\overline{Z}$ was 0.315. Our bootstrapped bandwidth is in the range of standard bandwidths used in two-dimensional density estimation (see, for example, Table 4.1 of Silverman, 1986).

Figure 1: Plot of the within residual's empirical cdf versus a normal cdf.

scale in the provision of capacity service is accepted at conventional significance levels using the within estimates, a finding in keeping with the bulk of empirical work in this industry, and is imposed in the three sets of estimates. Average annual productivity growth over the sample was a relatively healthy 2.2-2.3% for these firms. Owing to their clear size disadvantage, fleets of turboprop aircraft were unable to provide the route capacity enjoyed by jet aircraft. A 10% increase in density (ten percent reduction in number of route kilometers flown) would tend to increase the provision of route capacity by about 1.6 percent. The model fits for the three estimators are quite comparable. A Hausman-Wu specification test on the orthogonality conditions embedded in our efficient semiparametric (versus those of the within estimator) yielded a χ^2-statistic of 11.67 which is significant at only the 11% level and indicates that the orthogonality of firm efficiency levels with variables other than long-run movements in the labor force cannot be rejected at conventional significance levels. Finally, Table 1 points to 17-33% efficiency gains of the semiparametric efficient estimator over the efficient instrumental variables estimator of Hausman-Taylor. Table 2 presents the firm specific relative efficiency levels [(see Remark 3.1)] and rankings based on the three estimators[11]. The levels and rankings vary only marginally and point to an efficiency gap of 10 to 15% be-

[11]Standard errors for the relative ineffiiciencies are not available because they involve the asymptotic distribution of the difference between the random inefficiency and the maximum of the random inefficiency terms over the N firms in the sample. Such an asymptotic distribution has not yet been developed. We can, however, provide measures of samplying variation of $(\widehat{\alpha}_i - \widehat{\alpha}_j)$ and calculate this variation conditional on a particular selection for $\widehat{\alpha}_j$ − −the maximum of the $\widehat{\alpha}_i$ $(i = 1, ..., N)$. Based on estimates from the within, Hausman-Taylor, and the semiparametric efficient estimator, the standard errors of these differences are ??? 0.0253, 0.0262, 0.0233???.

tween the European and American carriers. The lowest ranked European carriers are Alitalia, Iberia, and SAS and they have been hemorrhaging financially during the 1990's. Of the lowest ranked American firms, Eastern has left the industry, and Continental emerged from bankruptcy just after the sample period ended.

5 Conclusions

This paper has introduced a new set of semiparametric efficient estimators which can be used to model random effects in panel data models. If all the regressors are allowed to be correlated with the effects (**Model 1**) then we have shown that the semiparametric efficient estimator is the traditional *within* estimator. When information is available that the effects are orthogonal to selected regressors (**Models 2 and 3**) *within* is no longer semiparametric efficient and we have derived the estimator under two different assumptions about short-run/long-run correlations between selected regressors and the effects. We have illustrated one of our new estimators (**Model 3**) by considering a newly constructed panel of European airlines and their American competitors to ascertain efficiency differentials in the provision of capacity service while allowing for inefficiency to be correlated with long-run adjustments in the personnel of national flag carriers. Often the national flag carriers are viewed by their governments as an extension of government and, as such, the potential for labor slack is quite real. Efficiency differences between the European and American carriers in our sample point to a gap of approximately 15% over the sample period 1976-90.

References

[1] Amemiya, T., and T.E. McCurdy (1986), "Instrumental-variable estimation of an error-components model", *Econometrica*, 54, 869-881.

[2] Bauer, P. (1990) "A survey of recent econometric developments in frontier estimation," *Journal of Econometrics,* 46, 21-39.

[3] 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.

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

[5] Breusch, T.S., Mizon, G.E., and P. Schmidt (1989), "Efficient estimation using panel data", *Econometrica*, 57, 695-700.

[6] 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.

[7] Good, D., Nadiri, M.I., Roller, Lars-Hendrik, and Sickles, R.C. (1993), Efficiency and productivity growth comparisons of European and U.S. air carriers: a first look at the data. *Journal of Productivity Analysis* 4, special issue edited by J. Mairesse and Z. Griliches, 115-125.

[8] Greene, W. H. (1996) Frontier production functions, *in Handbook of Applied Econometrics, Vol. II-Microeconometrics*, ed., H. Pesaran and P. Schmidt, Basil Blackwell.

[9] Hardy, G.H., Littlewood, J.E., and Polya, G. (1952), *Inequalities.* 2nd ed., Cambridge University Press.

[10] Hausman, J.A. and Taylor, W.E. (1981), Panel data and unobservable individual effects. *Econometrica*, 49, 1377-1398.

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

[12] 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.

[13] Newey, W.K. (1990), Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5, 99–136.

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

[15] Park, B.U. and Simar, L. (1994), Efficient semiparametric estimation in stochastic frontier models. *Journal of the American Statistical Association*, 89(427), 929–936.

[16] Robinson, P. (1988), Root-N-consistent semiparametric regression. *Econometrica 56*, 931–954.

[17] Roeller, L.H., and R.C. Sickles (1994), "Competition, market niches, and efficiency: a structural model of the European airline industry", (Rice University, mimeo).

[18] Schmidt, P. (1985) "Frontier production functions," *Econometrics Reviews*, 4, 289-328.

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

[20] Sickles, R. C. (1985) A nonlinear multivariate error components analysis of technology and specific factor productivity growth with an application to the U.S.airlines, *Journal of Econometrics*, 27, 61-78.

[21] Silverman, B.W. (1986), *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Table 1

Within, Efficient IV, and Efficient Semiparametric Estimates and Standard Errors

	Within		Efficient I.V.	
Variable	Parameter	Standard Error	Parameter	Standard
lnCapital	0.2952	0.0526	0.3341	0.0477
lnLabor	0.5204	0.0543	0.4943	0.0507
lnMaterials	0.1844	0.0349	0.1716	0.0345
Percent Turbo	−0.4979	0.1336	−0.4495	0.1282
Percent Wide	0.0074	0.1216	0.1065	0.1148
lnStagelength	0.1947	0.0751	0.2687	0.0697
lnNetworksize	−0.1600	0.0277	−0.1625	0.0259
lnLoadfactor	1.5351	0.1417	1.5286	0.1408
time	0.0230	0.00262	0.02220	0.00256
Adjusted R^2	0.8710		0.8437	

Table 1 (continued)

Efficient Semiparametric[12]

Variable	Parameter	Standard Error
lnCapital	0.3353	0.0395
lnLabor	0.4939	0.0408
lnMaterials	0.1708	0.02618
Percent Turbo	−0.4459	0.1003
Percent Wide	0.1122	0.0913
lnStagelength	0.2727	0.0563
lnNetworksize	−0.1624	0.0208
lnLoadfactor	1.5282	0.1064
time	0.02216	0.00197
Adjusted R^2	0.8452	

[12]The efficient instrumental variables estimator of Hausman and Taylor is used as the preliminary estimator to form the efficient semiparametric estimates.

Table 2

Relative Efficiency Levels and Rankings of European and American Airlines in the Provision of Capacity Service (1976-1990)

	Within		Hausman-Taylor		Efficient Semiparametric	
	Level	Rank	Level	Rank	Level	Rank
European Carriers						
Air France	0.799	5	0.816	5	0.815	5
Alitalia	0.663	13	0.696	13	0.698	13
British Air	0.738	9	0.766	9	0.766	9
Iberia	0.565	14	0.609	14	0.611	14
KLM	0.851	2	0.862	2	0.861	2
Lufthansa	0.764	8	0.806	7	0.808	7
SAS	0.546	15	0.589	15	0.591	15
Sabena	0.716	12	0.754	12	0.755	12
American Carriers						
American	0.765	7	0.784	8	0.785	8
Continental	0.737	10	0.756	11	0.757	11
Delta	0.767	6	0.812	6	0.814	6
Eastern	0.727	11	0.760	10	0.762	10
Northwest	1.000	1	1.000	1	1.000	1
TWA	0.828	3	0.845	3	0.846	3
United	0.826	4	0.844	4	0.845	4

Appendix 1

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. Throughout this section the time period T is considered fixed.

Write (X, Z, Y) for "generic" observations. Let $\mathbb{P}$ be the set of all possible joint distributions of (X, Z, Y). Let $\theta = (\beta', \gamma)'$. One calls $\mathbb{P}_0$ a regular parametric submodel of $\mathbb{P}$ if $\mathbb{P}_0(\subset \mathbb{P})$ can be represented as $\{P_{(\theta,\eta)} : \theta \in \mathbb{R}^d, \eta \in S$ 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).

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 $L(X, Z, Y,\ \theta, \eta)$ denote the log likelihood of an observation from $P_{(\theta,\eta)}$ and let $\ell_\theta(X, Z, Y) = \partial\ell/\partial\theta|_{(\theta_0,\eta_0)}$ and $\ell_{\eta_j}(X, Z, Y) = \partial\ell/\partial\eta_j|_{(\theta_0,\eta_0)}$ 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

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

where

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

$[\ell_\eta]$ 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 θ is called efficient if it is asymptotically normal with mean zero and variance $N^{-1} I^{-1}(P; \theta, \mathbb{P}_0)$.

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

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

A method of finding $I(P; \theta, \mathbb{P})$ 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 $I(P; \theta, \mathbb{P})$ can be obtained by

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

where

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

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.*

Appendix 2

Proof of Theorem 2.1.
With the notations of section 2.1., the log likelihood of (X, Y) can be written:

$$\begin{aligned} L(X, Y; \theta, \sigma^2, h) &= -\frac{T}{2}\log(2\pi\sigma^2) - \frac{\sum_{t=1}^{T}(Y_t - X_t'\theta)^2}{2\sigma^2} + \frac{(\overline{Y} - \overline{X}'\theta)^2}{2\overline{\sigma}^2} \\ &\quad + \log \int \exp\{-\frac{(\overline{Y} - \overline{X}'\theta - u)^2}{2\overline{\sigma}^2}\} h(u, X) du \end{aligned}$$

Let

$$w(s, x) = \int \phi_{\overline{\sigma}}(s - u) h(u, x) du \equiv \phi_{\overline{\sigma}} * h(\cdot, x)(s)$$

be the joint density of $(\overline{S}(\theta), X)$ on $\mathbb{R}^{1+Td}$, and $w'(s, x) = \frac{\partial}{\partial s} w(s, x)$
Then,

$$\begin{aligned} \ell_\theta = \frac{\partial L}{\partial \theta} &= \sigma^{-2} \sum_{t=1}^{T} U_t(\theta) X_t - \overline{X} \frac{w'}{w}(\overline{S}(\theta), X) \\ \ell_{\sigma^2} = \frac{\partial L}{\partial \sigma^2} &= (2\sigma^2)^{-1} \{\sum_{t=1}^{T} U_t^2(\theta)/\sigma^2 + \overline{\sigma}^{-1} \int [\overline{\sigma}^{-2}(\overline{S}(\theta) - u)^2 - T] \\ &\quad \cdot \phi(\overline{\sigma}^{-1}(\overline{S}(\theta) - u)) h(u, X) du / w(\overline{S}(\theta), X)\} \end{aligned}$$

Now

$$V = V_1 + V_2$$

where

$$V_1 = \{a(\overline{S}, X); a \in L_2(p), Ea(\overline{S}, X) = 0\}$$
$$V_2 = [\ell_{\sigma^2}]$$

Lemma A.1.

$$E\left[\frac{w'}{w}(\overline{S}(\theta), X) \mid X\right] = 0$$

Proof: Note first that conditional pdf of $\overline{S}(\theta)$ given $X = x$ is given by

$$\int \phi_{\overline{\sigma}}(s-u)h(u|x)du$$

where $h(u|x)$ denotes the conditional pdf of α given $X = x$. We can write

$$\begin{aligned}
&E\left[\frac{w'}{w}(\overline{S}(\theta), X) \mid X\right] \\
&= \int \left\{ \frac{\frac{\partial}{\partial s}\int \phi_{\overline{\sigma}}(s-u)h(u,X)du}{\int \phi_{\overline{\sigma}}(s-u)h(u,X)du} \int \phi_{\overline{\sigma}}(s-u)h(u|x)du \right\} ds \\
&= \int \frac{\partial}{\partial s}\{ \int \phi_{\overline{\sigma}}(s-u)h(u|X)du\} ds \\
&= \frac{\partial}{\partial s}[\int \{\int \phi_{\overline{\sigma}}(s-u)h(u|X)du\} ds] \\
&= \frac{\partial}{\partial s}[1] \\
&= 0
\end{aligned}$$

Lemma A.2.

$$E(U_t(\theta) \mid \overline{S}(\theta), X) = 0$$

Proof: First, note that $(\varepsilon_1 - \overline{\varepsilon}, \ldots, \varepsilon_T - \overline{\varepsilon})$ is independent of $\overline{\varepsilon}$ since $\varepsilon_i \sim IID\ N(0, \sigma^2)$, and of course is independent of α and X too. This means that $U_t(\theta) \equiv \varepsilon_t - \overline{\varepsilon}$ is independent of $\overline{S}(\theta) \equiv \alpha + \overline{\varepsilon}$ and X. Hence,

$$E(U_t(\theta) \mid \overline{S}(\theta), X) = E(U_t(\theta)) = 0$$

Now, from Lemma A.1. and Lemma A.2., we get $E(\ell_\beta) = 0$. Thus,

$$\begin{aligned}\pi(\ell_\beta \mid V_1) &= E(\ell_\beta \mid \overline{S}, X) - E(\ell_\beta) \\ &= -\overline{X}\frac{w'}{w}(\overline{S}, X)\end{aligned}$$

Hence,

$$\begin{aligned}\ell^* &= \ell_\beta - \pi(\ell_\beta | V_1) - \pi(\ell_\beta - \pi(\ell_\beta \mid V_1) \mid W) \\ &= \sigma^{-2}\sum_{t=1}^{T} U_t(\theta)X_t - \pi(\sigma^{-2}\sum_{t=1}^{T} U_t(\theta)X_t \mid W)\end{aligned}$$

where

$$W = [\ell_{\sigma^2} - \pi(\ell_{\sigma^2} \mid V_1)]$$

Now, one can write

$$\ell_{\sigma^2} - \pi(\ell_{\sigma^2} \mid V_1) = (2\sigma^4)^{-1}\sum_{t=1}^{T} U_t^2(\theta) + c(\overline{S}, X)$$

for some function c. Note that

$$\begin{aligned}E(\sum_{t=1}^{T} U_t(\theta)X_t \cdot c(\overline{S}, X)) &= E[E(\sum_{t=1}^{T} U_t(\theta)X_t \cdot c(\overline{S}, X) \mid \overline{S}, X)] \qquad \text{(A.1)} \\ &= E[\sum_{t=1}^{T} X_t c(\overline{S}, X) E(U_t(\theta) \mid \overline{S}, X)] \\ &= 0\end{aligned}$$

by lemma A.2. And,

$$E(\sum_{t=1}^{T} U_t(\theta)X_t \cdot \sum_{t=1}^{T} U_t^2(\theta)) = -E(\sum_{t=1}^{T} U_t(\theta)X_t \cdot \sum_{t=1}^{T} U_t^2(\theta))$$

$$\begin{aligned}\text{since } (U_1(\theta), \ldots, U_T(\theta)) &= (\varepsilon_1 - \overline{\varepsilon}, \ldots, \varepsilon_T - \overline{\varepsilon}) \\ &\overset{\mathrm{d}}{=} -(\varepsilon_1 - \overline{\varepsilon}, \ldots, \varepsilon_T - \overline{\varepsilon}) \\ &= -(U_1(\theta), \ldots, U_T(\theta))\end{aligned}$$

Here, "$\overset{\mathrm{d}}{=}$" means the distributions are the same. Hence

$$E(\sum_{t=1}^{T} U_t(\theta)X_t \cdot \sum_{t=1}^{T} U_t^2(\theta)) = 0 \qquad \text{(A.2)}$$

The results (A.1) and (A.2) imply that

$$\sigma^{-2} \sum_{t=1}^{T} U_t(\theta) X_t \perp W$$

or equivalently

$$\pi(\sigma^{-2} \sum_{t=1}^{T} U_t(\theta) X_t \mid W) = 0$$

Therefore

$$\ell^* = \sigma^{-2} \sum_{t=1}^{T} U_t(\theta) X_t$$

And the information is

$$\begin{aligned}
I(P; \theta, \mathbb{P}) &= E\ell^* \ell^{*'} \\
&= \sigma^{-4} E(\sum_{t=1}^{T} (\varepsilon_t - \overline{\varepsilon}) X_t)(\sum_{t=1}^{T} (\varepsilon_t - \overline{\varepsilon}) X_t') \\
&= \sigma^{-4} E(\sum_{t=1}^{T} \varepsilon_t (X_t - \overline{X}))(\sum_{t=1}^{T} \varepsilon_t (X_t - \overline{X})') \\
&= \sigma^{-4} E(\sum_{t=1}^{T} \varepsilon_t^2 (X_t - \overline{X})(X_t - \overline{X})') \\
&= \sigma^{-2} T E(\frac{1}{T} \sum_{t=1}^{T} (X_t - \overline{X})(X_t - \overline{X})') \\
&\equiv \overline{\sigma}^{-2} \Sigma_W
\end{aligned}$$

Proof of Theorem 2.2.
With the notations of section 2.2., we have the log likelihood of (X, Y, Z):

$$\begin{aligned}
&L(X, Z, Y; \beta, \gamma, \sigma^2, h, g) \\
&\quad = \log g(X \mid Z) - \frac{T}{2} \log(2\pi\sigma^2) \\
&\qquad - \frac{\sum_{t=1}^{T} (Y_t - X_t'\beta - Z_t'\gamma)^2}{2\sigma^2} + \frac{(\overline{Y} - \overline{X}'\beta - \overline{Z}'\gamma)^2}{2\overline{\sigma}^2} \\
&\qquad + \log \int \exp[-\frac{(\overline{Y} - \overline{X}'\beta - \overline{Z}'\gamma - u)^2}{2\overline{\sigma}^2}] h(u, Z) du
\end{aligned}$$

Then,

$$\ell_\beta = \frac{\partial L}{\partial \beta} = \sigma^{-2}\sum_{t=1}^{T} U_t(\theta)X_t - \overline{X}\frac{w'}{w}(\overline{S}(\theta), Z)$$

$$\ell_\gamma = \frac{\partial L}{\partial \gamma} = \sigma^{-2}\sum_{t=1}^{T} U_t(\theta)Z_t - \overline{Z}\frac{w'}{w}(\overline{S}(\theta), Z)$$

$$\begin{aligned}\ell_{\sigma^2} = \frac{\partial L}{\partial \sigma^2} &= (2\sigma^2)^{-1}\{\sum_{t=1}^{T} U_t^2(\theta)/\sigma^2 + \overline{\sigma}^{-1}\int [\overline{\sigma}^{-2}(\overline{S}(\theta)-u)^2 - T] \\ &\cdot\phi(\overline{\sigma}^{-1}(\overline{S}(\theta)-u))h(u,Z)du/w(\overline{S}(\theta), Z)\}\end{aligned}$$

In this model,

$$V = V_1 + V_2 + V_3$$

where

$$\begin{aligned}&V_1 = \{a(X,Z); a \in L_2(p), \quad E(a(X,Z) \mid Z) = 0\}\\ &V_2 = \{b(\overline{S},Z); b \in L_2(p), \quad E(b(\overline{S},Z)) = 0\}\\ &V_3 = [\ell_{\sigma^2}]\end{aligned}$$

Lemma A.3.

$$E[\frac{w'}{w}(\overline{S}(\theta), Z) \mid X, Z] = 0$$

Proof: Since X and α are independent conditionally on Z,

$$E\left[\frac{w'}{w}(\overline{S}(\theta), Z) \mid X, Z\right] = E\left[\frac{w'}{w}(\overline{S}(\theta), Z) \mid Z\right]$$

And the rest of the proof is the same as that of Lemma A.1.

Lemma A.4.

$$E(U_t(\theta) \mid \overline{S}(\theta), X, Z) = 0$$

Proof: The same as Lemma A.2.

Note

$$\ell_\beta^* = \ell_\beta - \pi(\ell_\beta \mid V_1 + V_2) - \pi(\ell_\beta - \pi(\ell_\beta \mid V_1 + V_2) \mid W_1) \tag{A.3}$$

where

$$W_1 = [\ell_{\sigma^2} - \pi(\ell_{\sigma^2} \mid V_1 + V_2)]$$

Furthermore,

$$\begin{aligned}\ell_\beta - \pi(\ell_\beta \mid V_1 + V_2) \quad &= \quad \ell_\beta - \pi(\ell_\beta \mid V_1) \\ &\qquad -\pi(\ell_\beta - \pi(\ell_\beta \mid V_1) \mid W_2)\end{aligned} \tag{A.4}$$

where

$$\begin{aligned} W_2 &= V_1^\perp \cap (V_1 + V_2) \\ &= \{b(\overline{S}, Z) - E(b(\overline{S}, Z) \mid Z); b \in L_2(p)\} \end{aligned}$$

By Lemma A.3. and Lemma A.4., $\ell_\beta \perp V_1$ hence

$$\pi(\ell_\beta \mid V_1) = 0 \tag{A.5}$$

Hence by (A.3), (A.4) and (A.5)

$$\ell_\beta^* = \ell_\beta - \pi(\ell_\beta \mid W_2) - \pi(\ell_\beta - \pi(\ell_\beta \mid W_2) \mid W_1) \tag{A.6}$$

From Lemma A.4.

$$\sigma^{-2} \sum_{t=1}^{T} U_t(\theta) X_t \perp W_2$$

Thus,

$$\begin{aligned} \pi(\ell_\beta \mid W_2) &= \pi(-\overline{X}\frac{w'}{w}(\overline{S}, Z) \mid W_2) \\ &= E(-\overline{X}\frac{w'}{w}(\overline{S}, Z) \mid \overline{S}, Z) \\ &\quad - E(-\overline{X}\frac{w'}{w}(\overline{S}, Z) \mid Z) \end{aligned}$$

Since X and α are independent conditionally on Z,

$$\begin{aligned} E(-\overline{X}\frac{w'}{w}(\overline{S}, Z) \mid \overline{S}, Z) &= E(-\overline{X} \mid \overline{S}, Z)\frac{w'}{w}(\overline{S}, Z) \\ &= -E(\overline{X} \mid Z)\frac{w'}{w}(\overline{S}, Z) \end{aligned}$$

and

$$\begin{aligned} E(-\overline{X}\frac{w'}{w}(\overline{S}, Z) \mid Z) &= E(-\overline{X} \mid Z)E(\frac{w'}{w}(\overline{S}, Z) \mid Z) \\ &= 0 \end{aligned}$$

by Lemma A.3. These imply

$$\pi(\ell_\beta \mid W_2) = -E(\overline{X} \mid Z)\frac{w'}{w}(\overline{S}, Z)$$

Therefore

$$\begin{aligned} \ell_\beta - \pi(\ell_\beta \mid W_2) &= \sigma^{-2} \sum_{t=1}^{T} U_t(\theta) X_t \\ &\quad -(\overline{X} - E(\overline{X} \mid Z))\frac{w'}{w}(\overline{S}(\theta), Z) \end{aligned} \tag{A.7}$$

Now, observe that

$$\ell_{\sigma^2} - \pi(\ell_{\sigma^2} \mid V_1 + V_2) = (2\sigma^4)^{-1} \sum_{t=1}^{T} [U_t^2(\theta) - E(U_t^2(\theta))]$$

which is perpendicular to $\ell_\beta - \pi(\ell_\beta \mid W_2)$. Thus from (A.6) and (A.7) we get

$$\ell_\beta^* = \sigma^{-2} \sum_{t=1}^{T} U_t(\theta) X_t - (\overline{X} - E(\overline{X} \mid Z)) \frac{w'}{w} (\overline{S}(\theta), Z)$$

Parallel arguments leading to ℓ_β^* can be applied to ℓ_γ, which yields

$$\begin{aligned} \ell_\gamma^* &= \sigma^{-2} \sum_{t=1}^{T} U_t(\theta) Z_t - (\overline{Z} - E(\overline{Z} \mid Z)) \frac{w'}{w} (\overline{S}(\theta), Z) \\ &= \sigma^{-2} \sum_{t=1}^{T} U_t(\theta) Z_t \end{aligned}$$

The information is now

$$\begin{aligned} I(P; \theta, \mathbb{P}) &= E\ell^* \ell^{*'} \\ &= \begin{pmatrix} \overline{\sigma}^{-2} \Sigma_W(X) + I_0 \Sigma_B(X|Z) & \overline{\sigma}^{-2} \Sigma_W(X, Z) \\ \overline{\sigma}^{-2} \Sigma_W(X, Z)' & \overline{\sigma}^{-2} \Sigma_W(Z) \end{pmatrix} \end{aligned}$$

Proof of Theorem 2.3. Let $f_1 * f_2$ denotes the convolution of $1+Tq$ dimensional density functions f_1 and f_2, i.e., for $x_1 \in R'$ and $x_2 \in R^{tq}$,

$$f_1 * f_2(x_1, x_2) = \int f_1(x_1 - y_1, x_2 - y_2) f_2(y_1, y_2) dy_1 dy_2.$$

Then we can write

$$\begin{aligned} w_N(s, z) &= E\hat{w}(s, z) = L_{s_N} * w(s, z) + c_N \\ &= \int \phi_{\overline{\sigma}}(s - u) h_N(u, z) du + c_N \end{aligned} \qquad (A.8)$$

where $h_N = h * L_{s_N}$ and $L_{s_N}(s, z) = K_{s_N}(s) \prod_{t=1}^{T} \prod_{k=1}^{q} K_{s_N}(z_{tk})$. We begin by showing

$$E\left[\frac{w_N'}{w_N}(\overline{S}(\theta), Z) - \frac{w'}{w}(\overline{S}(\theta), Z))\right]^2 \to 0. \qquad (A.9)$$

From (A.8), we can write

$$w_N'(s,z) = \overline{\sigma}^{-3} \int (u-s)\phi[\overline{\sigma}^{-1}(s-u)]h_N(u,z)du.$$

The absolute value of the above term is now bounded by the term

$$\overline{\sigma}^{-2} \int |u-s|h_N(u,z)du \cdot w_N(s,z)$$

due to an inequality of Chebyshev [Hardy, Littlewood and Pólya (1952), page 43]. Hence

$$\left(\frac{w_N'}{w_N}(s,z)\right)^2 \leq 2\overline{\sigma}^{-4}[\int |u|^2 h_N^2(u,z)du + \int |s|^2 h_N^2(u,z)du].$$

The fact that $w_N'/w_N(s,z) \to w'/w(s,z)$ and $h_N(s,z) \to h(s,z)$ for every (s,z) reduces, by dominated convergence, the proof of (A.9) to showing

$$E\int |u|^2 h_N^2(u,Z)du \to E\int |u|^2 h^2(u,\overline{Z})du \tag{A.10}$$

$$E\int |\overline{S}|^2 h_N^2(u,Z)du \to E\int |\overline{S}|^2 h^2(u,Z)du \tag{A.11}$$

For (A.10), by exchanging the order of integrations,

$$E\int |u|^2 h_N^2(u,Z)du = \int\int [\int |u|^2 h_N^2(u,z)w(s,z)ds]dudz.$$

Since h_N and the marginal densities of w are bounded, the integral inside the bracket is bounded by $C|u|^2 h_N(u,z)$. Furthermore,

$$\int\int |u|^2 h_N(u,z)dudz = E|\alpha|^2 + O(s_N^2)$$

$$\to E|\alpha|^2 = \int\int |u|^2 h(u,z)dudz.$$

By dominated convergence, (A.10) is verified. Finally, (A.11) is an immediate consequence of boundedness of h_N and dominated convergence.

Now, note that the theorem is implied by

$$N^{\frac{1}{2}}\left[\hat{\theta}_N - \left(\theta + I_N^{-1}N^{-\frac{1}{2}}Q_N(\theta)\right)\right] \xrightarrow{p} 0 \tag{A.12}$$

$$I_N \xrightarrow{p} I \tag{A.13}$$

$$N^{-\frac{1}{2}}\sum_{i=1}^{N}(\hat{p}(Z_i) - p(Z_i))\left[\frac{w_N'}{w_N}(\overline{S}_i(\theta), Z_i)\right] \xrightarrow{p} 0 \tag{A.14}$$

where

$$
Q_N(\theta) = N^{-\frac{1}{2}} \begin{pmatrix} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} U_{it}(\theta) X_{it}/\sigma^2 - (\overline{X}_i - \hat{p}(Z_i)) \frac{w'_N}{w_N} (\overline{S}_i(\theta), Z_i) \right] \\ \\ \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}(\theta) Z_{it}/\sigma^2 \end{pmatrix}
$$

$$
I_N = \begin{pmatrix} \overline{\sigma}^{-2} \sum_W(X) + I_{0N}\Sigma_B(X|Z) & \overline{\sigma}^{-2}\Sigma_W(X,Z) \\ \\ \overline{\sigma}^{-2}\Sigma_W(X,Z)' & \overline{\sigma}^{-2} \sum_W(Z) \end{pmatrix}
$$

$$
I_{0N} = \int (w'_N/w_N)^2 w
$$

The convergence (A.13) is immediate consequence of (A.9). For (A.14), note that the left hand side is equal to

$$
(\widehat{a} - a)N^{-1/2} \sum_{i=1}^{N} \left[\frac{w'_N}{w_N} (\overline{S}_i(\theta), Z_i) \right] + (\hat{A} - A)N^{-1/2} \sum_{i=1}^{N} Z_i \left[\frac{w'_N}{w_N} (\overline{S}_i(\theta), Z_i) \right]
$$

where a, $\widehat{a}$, A, and $\widehat{\mathrm{A}}$ are defined in section 2.2. Again, by (A.9), the above two terms goes to zero in probability as $N \to \infty$. Now (A.12) is implied by

$$
\sup \mid Q_N(\theta') - Q_N(\theta) + I_N N^{\frac{1}{2}}(\theta' - \theta) \mid \xrightarrow{p} 0
$$

$$
\sup \mid N^{-\frac{1}{2}} \sum_{i=1}^{N} \{ \frac{\hat{w}'}{\hat{w}} \big(\overline{S}_i(\theta'), Z_i, \theta'\big) - \frac{w'_N}{w_N} \big(\overline{S}_i(\theta'), Z_i\big) \} (\overline{X}_i - \hat{p}(Z_i)) \mid \xrightarrow{p} 0
$$

$$
\sup \mid \hat{\sigma}^2(\theta') - \hat{\sigma}^2(\theta) \mid \xrightarrow{p} 0
$$

$$
\hat{I} \xrightarrow{p} I
$$

$$
N^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}(\tilde{\theta}_N) \begin{pmatrix} X_{it} \\ Z_{it} \end{pmatrix} = O_p(1)
$$

where all“sup” mean supremum over all θ' such that $N^{\frac{1}{2}} \mid \theta' - \theta \mid \leq M$ for some $M > 0$. The proofs of all of these can be carried out in similar fashion as in Park and Simar (1994). Utilizing these facts sample splitting is not necessary to avoid complicated asymptotic arguments such as uniform convergence.

Proof of Theorem 2.4. For simplicity drop the subscript (N,T) in $\tilde{\theta}_{N,T}, \hat{\theta}_{N,T}$ and $s_{N,T}$. Note that one can write

$$\begin{aligned}\hat{\theta}-\theta = (\tilde{\theta}-\theta) + N^{-1}\hat{I}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}(\tilde{\theta})R_{it}/\hat{\sigma}^2(\tilde{\theta}) \\ -N^{-1}\hat{I}^{-1}\begin{pmatrix}\sum_{i=1}^{N}(\overline{X}_i-\hat{p}(Z_i))\frac{\hat{w}'}{\hat{w}}(\overline{S}_i(\tilde{\theta}),Z_i,\tilde{\theta}) \\ 0\end{pmatrix} \\ = A_1 + A_2 + A_3\end{aligned}$$

Since $\hat{\sigma}^2(\tilde{\theta}) = O_p(1)$ we have $\hat{I} = O_p(T)$, hence $|A_3|$ is bounded by $O_p(T^{-1}s^{-1})$. Observe that

$$\sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}(\tilde{\theta})R_{it} = -NT\hat{\Sigma}_W(\tilde{\theta}-\theta) + \sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}(\theta)R_{it} \tag{A.15}$$

Noting that $T\hat{I}^{-1}\hat{\Sigma}_W/\hat{\sigma}^2(\tilde{\theta}) = J + O_p(T^{-1}s^{-2})$ where J is the identity matrix, we have by (A.15)

$$A_1 + A_2 = (\tilde{\theta}-\theta)O_p(T^{-1}s^{-2}) + N^{-1}\hat{I}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}(\theta)R_{it}/\hat{\sigma}^2(\tilde{\theta}) \tag{A.16}$$

Combining (A.16), (2.7) and the fact that $(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}(\theta)R_{it} = O_p(1)$, we get $(NT)^{1/2}(A_1+A_2) = O_p(1)$. This establishes *(ii)*.

The proof of *(i)* is virtually the same, hence omitted.

Proof of Theorem 3.1. Since $\alpha_i = \overline{S}_i(\theta) - \overline{\varepsilon}_i$ and $\hat{\alpha}_i = \overline{S}_i(\hat{\theta})$ we have:

$$T^{1/2}(\hat{\alpha}_i - \alpha_i) = -T^{1/2}\overline{R}_i'(\hat{\theta}-\theta) + T^{1/2}(\overline{S}_i(\theta)-\alpha_i)$$

The first term tends to zero by Theorem 2.4. The second term is $T^{1/2}\overline{\varepsilon}_i$ which has the $N(0,\sigma^2)$ distribution.

Proof of Theorem 3.2. Case *(i)* is obvious from Theorem 3.1. Case *(ii)* follows since $E\max_{1\le i\le N}|\overline{X}_i|, E\max_{1\le i\le N}|\overline{Z}_i|$ and $E\max_{1\le i\le N}|T^{1/2}\overline{\varepsilon}_i|$ are $O(\log N)$.

Proof of Theorem 3.3. For any given $M > 0$,

$$P[N^{\frac{1}{\delta+1}}|\alpha_{(N)} - B(h)| \le M] = P[\alpha_{(N)} < B(h) - MN^{-\frac{1}{\delta+1}}]$$

$$= [1 - h(B(h) - \delta M N^{-\frac{1}{\delta+1}}) M N^{-\frac{1}{\delta+1}}]^N$$

where $0 \leq \delta \leq 1$. Using condition (3.9), we can bound the probability by $(1 - D_1 M^{\delta+1} N^{-1})^N$ for some $D_1 > 0$. Since this can be bounded by $\exp(-D_2 M^{\delta+1})$ for some $D_2 > 0$, the theorem follows.