Semiparametric Efficient Estimation of Dynamic 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

May 19, 2003

Abstract

This paper extends the semiparametric efficient treatment of panel data models pursued by Park and Simar (1994) and Park, Sickles, and Simar (1998, 2003) to a dynamic panel setting. We develop a semiparametric efficient estimator under minimal assumptions when the panel model contains a lagged dependent variable. We apply this new estimator to analyze the structure of demand between city pairs for selected U. S. airlines during the period 1979 I to 1992 IV.

*Research of B. U. Park was supported by the Korea Research Foundation Grant KRF-2002-070-C00017.

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

1 Introduction

Arellano and Bond (1991), Arellano and Bover (1995), and Ahn and Schmidt (1995) address the question of efficient estimation in dynamic panel models by investigating the number of moment conditions available under several sets of assumptions about the relationship between the initial condition and the error terms, building on earlier work by Anderson and Hsiao (1981, 1982). Once these moment conditions have been identified, the generalized method of moments (GMM) technique can be applied to obtain efficient estimates, utilizing the moment conditions described by Ahn and Schmidt, as well as those implied by exogeneity assumptions on the other regressors in the model. The estimates are efficient as long as the correct moment conditions are specified. A number of excellent surveys and monographs have been written on the subject, mostly recently by Baltagi (1995) and Mátyás and Sevestre (1996). These authors discuss a number of alternative estimators to be applied to random effects models of the form we consider herein. The key differences among the various estimators of the dynamic panel data model essentially involve the imposition of different orthogonality conditions to yield different sets of instruments. Which estimator is better in the sense of a smaller asymptotic variance is difficult to analyze. The class of GMM estimators which are efficient (Ahn and Schmidt, 1995; Arellano and Bovwer, 1995) have been shown to be difficult to implement in large data sets.

Building on previous work of Park and Simar (1994), and Park, Sickles, and Simar (1998, 2003) our paper utilizes a somewhat different approach than that of Ahn and Schmidt and instead of finding the orthogonality conditions necessary to achieve the efficiency bound, constructs the estimator which attains the semiparametric efficiency bound. We analyze our estimator using Monte Carlo simulations. Our semiparametric efficient estimator is developed under minimal assumptions when the panel model contains a lagged dependent variable. We apply this new estimator in an analysis of the structure of dynamic demand in markets (city-pairs) for selected U. S. airlines during the period 1979 I to 1992 IV. Our results suggest that this estimator may have advantages over parametric estimators in regard to efficiency gains.

2 Main Results

The model we analyze in this paper is the dynamic panel data model that can be written as:

$$Y_{it} = \gamma Y_{i,t-1} + \beta' X_{it} + \alpha_i + \varepsilon_{it}\,; \quad i = 1, \ldots, n\,;\ t = 1, \ldots, r \tag{2.1}$$

where $X_{it} \in \mathbb{R}^d$, $\beta \in \mathbb{R}^d$ and ε_{it} are iid random variables from a $\mathcal{N}(0, \sigma^2)$ with an unknown σ^2. We assume $|\gamma| < 1$ and $Y_{i,0} = 0$. The random effects α_i are assumed to be independent and have an unknown common density function h. Write $\varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{ir})'$, $X_i \equiv (X'_{i1}, \ldots, X'_{ir})'$, and $Y_i \equiv (Y_{i1}, \ldots, Y_{ir})'$. The random covariates X_i are independent and identically distributed with an unknown density function g defined on $\mathbb{R}^{dr}$. It is assumed that ε's, α's and X's are independent. In this section we address efficient estimation of the parameters β and γ in the presence of the nuisance parameters σ^2, h and g. Note that the parameter spaces for h and g are of infinite dimension while those for β, γ and σ^2 are of finite dimension, so the model (2.1) is semiparametric.

We speak of efficiency as n tends to infinity with the time period r being fixed. The notion of efficiency in the semiparametric world is well explained in Bickel *et al.* (1993). There is a Fisher-like information matrix, say I, such that all *regular* estimators have asymptotic covariance matrices that are greater than or equal to I (Hájek-Le Cam's Convolution Theorem, see Theorem 2.3.1 of Bickel *et al.*, 1993). Here, we say an estimator $\hat{\delta}_n$ of $q(\theta)$ is regular if the law of $\sqrt{n}(\hat{\delta}_n - q(\theta_n))$ under P_{θ_n} converges to a limit law whenever $\sqrt{n}|\theta_n - \theta|$ stays bounded, and if the limit distribution does not depend on the choice of $\{\theta_n\}$. We call $\hat{\delta}_n$ *efficient* if its limit law is $\mathcal{N}(0, I^{-1})$. In the next subsection we exhibit the information matrix I for estimating $(\beta', \gamma)'$ in the presence of the nuisance parameters σ^2, $h(\cdot)$ and $g(\cdot)$, and then in the second subsection we construct an efficient estimator of $(\beta', \gamma)'$.

2.1 Information bound

Let $Z_{1t} \equiv Z_{1t}(\beta, \gamma) = Y_{1t} - \gamma Y_{1,t-1} - \beta' X_{1t}$ and $\bar{Z}_1 \equiv \bar{Z}_1(\beta, \gamma) = \sum_{t=1}^r Z_{1t}(\beta, \gamma)/r$. Define $\bar{\sigma}^2 = \sigma^2/r$. Then we can write $Z_{1t} = \alpha_1 + \varepsilon_{1t}$, $\bar{Z}_1 = \alpha_1 + \bar{\varepsilon}_1$ and $Z_{1t} - \bar{Z}_1 = \varepsilon_{1t} - \bar{\varepsilon}_1$. The probability density function for $\bar{Z}_1$ is given by

$$w(z) \equiv w(z; \sigma^2, h(\cdot)) = \int (2\pi\bar{\sigma}^2)^{-1/2} \exp\{-(z-u)^2/(2\bar{\sigma}^2)\} h(u)\, du.$$

Thus, the log-likelihood with a single observation (X_1, Y_1) is given by

$$\begin{aligned} L(\beta, \gamma, \sigma^2, h(\cdot), g(\cdot); X_1, Y_1) \quad &= \quad \log g(X_1) - \frac{r}{2}\log(2\pi\sigma^2) - \sum_{t=1}^r \frac{Z_{1t}^2}{2\sigma^2} + \frac{\bar{Z}_1^2}{2\bar{\sigma}^2} \\ &\quad + \log w(\bar{Z}_1) + \frac{1}{2}\log(2\pi\bar{\sigma}^2). \end{aligned} \tag{2.2}$$

Write $P_{\beta,\gamma,\sigma^2,h,g}$ for the probability distribution of (X_1, Y_1) corresponding to β, γ, σ^2, h and g. Let β_0, γ_0, σ_0^2, h_0 and g_0 be the true values and the true functions, thus the true probability distribution is $P_0 = P_{\beta_0,\gamma_0,\sigma_0^2,h_0,g_0}$. For the time being, let us suppose the model (2.1), denoted by $\mathcal{P}$, is parametric. Let $\mathcal{P} = \{P_{\beta,\gamma,\sigma^2,h(\cdot;\eta_1),g(\cdot;\eta_2)} \;:\; \beta \in \mathbb{R}^d, \gamma \in \mathbb{R}, \sigma^2 \in$

$\mathbb{R}^+, \eta_1 \in S_1, \eta_2 \in S_2\}$ for some open S_1, $S_2 \subset \mathbb{R}$ where $h(\cdot;\eta_1)$ and $g(\cdot;\eta_2)$ are known except η_1 and η_2. If the maps $\eta_1 \to h^{1/2}(\cdot;\eta_1)$ and $\eta_2 \to g^{1/2}(\cdot;\eta_2)$ from $\mathbb{R}$ to $L_2(\mu)$ (μ is the Lebesgue measure) are "smooth", then the model $\mathcal{P}$ is regular. (See Ibragimov & Has'minskii, 1981, Section 1.7, or Bickel *et al.*, 1993, Section 2.1). For this regular parametric model $\mathcal{P}$, the information matrix, denoted by $I(P_0 \,|\, \beta, \gamma, \mathcal{P})$, for estimating β and γ in the presence of the nuisance parameters σ^2, η_1 and η_2 is well defined and can be computed in the following way.

Write $L = L(\beta, \gamma, \sigma^2, h(\cdot;\eta_1), g(\cdot;\eta_2); X_1, Y_1)$ and define $\ell_\beta = \partial L/\partial\beta|_{\beta_0,\gamma_0,\sigma_0^2,\eta_{10},\eta_{20}}$ where η_{10} and η_{20} are the parameter values such that $h_0 = h(\cdot, \eta_{10})$ and $g_0 = g(\cdot, \eta_{20})$. Define ℓ_γ, ℓ_{σ^2}, ℓ_{η_1} and ℓ_{η_2}, likewise. Let $[\ell_{\sigma^2}, \ell_{\eta_1}, \ell_{\eta_2}]$ be the linear span generated by ℓ_{σ^2}, ℓ_{η_1} and ℓ_{η_2}. Define $\ell_\beta^* = \ell_\beta - \Pi(\ell_\beta \,|\, [\ell_{\sigma^2}, \ell_{\eta_1}, \ell_{\eta_2}])$, and likewise define ℓ_γ^*, where $\Pi(u \,|\, \mathcal{S})$ denotes the vector of projections of each component of u onto the space $\mathcal{S}$ in $L_2(\mu)$. Write $\ell^* = (\ell_\beta^{*\prime}, \ell_\gamma^*)'$. The information matrix is then given by

$$I(P_0 \,|\, \beta, \gamma, \mathcal{P}) = E_{P_0}\, \ell^* \ell^{*\prime}. \tag{2.3}$$

It is known that the right hand side of (2.3) equals the inverse matrix of the $(d+1)\times(d+1)$ left-top block of the matrix $\{E_{P_0}\, \ell\, \ell'\}^{-1}$ where $\ell = (\ell_\beta', \ell_\gamma, \ell_{\sigma^2}, \ell_{\eta_1}, \ell_{\eta_2})'$. It is also known that if δ is a Gaussian regular estimator of $(\beta_0', \gamma_0)'$ with asymptotic covariance $\Sigma(P_0, \delta)$ then

$$\Sigma(P_0, \delta) \geq I^{-1}(P_0 \,|\, \beta, \gamma, \mathcal{P}),$$

where $A \geq B$ for matrices A and B means that $A - B$ is nonnegative definite (Bickel *et al.*, 1993, Section 2.3).

Now we go back to the original semiparametric model where the spaces for h and g are of infinite dimension. Consider classes of functions $h(\cdot;\eta_1)$ and $g(\cdot;\eta_2)$ indexed by $\eta_1, \eta_2 \in \mathbb{R}$ such that $h(\cdot;0) = h_0$ and $g(\cdot;0) = g_0$. Form a parametric submodel $\mathcal{P}_0 = \{P_{\beta,\gamma,\sigma^2,h(\cdot;\eta_1),g(\cdot;\eta_2)} \,:\, \beta \in \mathbb{R}^d, \gamma \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+, \eta_1 \in \mathbb{R}, \eta_2 \in \mathbb{R}\}$. If we choose $h(\cdot;\cdot)$ and $g(\cdot;\cdot)$ so that the maps $\eta_1 \to h^{1/2}(\cdot;\eta_1)$ and $\eta_2 \to g^{1/2}(\cdot;\eta_2)$ from $\mathbb{R}$ to $L_2(\mu)$ are "smooth", then $\mathcal{P}_0$ is a regular parametric submodel of $\mathcal{P}$ and the information matrix $I(P_0 \,|\, \beta, \gamma, \mathcal{P}_0)$ can be defined as at (2.3). Consider the class of all such regular parametric submodels, and write it $\mathcal{C}$. Suppose an estimator δ of $(\beta_0', \gamma_0)'$ is Gaussian regular on $\mathcal{P}$. Then it is Gaussian regular on every regular parametric submodel $\mathcal{P}_0$, too. So, it satisfies

$$\Sigma(P_0, \delta) \geq I^{-1}(P_0 \,|\, \beta, \gamma, \mathcal{P}_0), \tag{2.4}$$

for every regular parametric submodel $\mathcal{P}_0$. In view of (2.4) it is natural to define the information bound for estimating $(\beta', \gamma)'$ in the semiparametric model by

$$I^{-1}(P_0 \,|\, \beta, \gamma, \mathcal{P}) = \sup\{I^{-1}(P_0 \,|\, \beta, \gamma, \mathcal{P}_0) \,:\, \mathcal{P}_0 \in \mathcal{C}\}. \tag{2.5}$$

A method of calculating $I(P_0 \mid \beta, \gamma, \mathcal{P})$ can be found in Bickel *el al.* (1993). The main tasks are to find the *tangent space* of $\mathcal{P}_{\rm nu} = \{P_{\beta_0,\gamma_0,\sigma^2,h,g} : \sigma^2 \in \mathbb{R}^+, \int h = 1, \int g = 1, h, g \geq 0\}$ at (σ_0^2, h_0, g_0), and to calculate the orthogonal projection of the scores ℓ_β and ℓ_γ onto the tangent space. Let $\ell_{\rm nu}(\mathcal{P}_0) = (\ell_{\sigma^2}, \ell_{\eta_1}, \ell_{\eta_2})$ be the vector of scores for the nuisance parameters σ^2, η_1 and η_2. We introduce $\mathcal{P}_0$ here to stress its dependence on the choice of parametric submodel $\mathcal{P}_0$. Then, the tangent space of $\mathcal{P}_{\rm nu}$ at (σ_0^2, h_0, g_0) is nothing else than the closed linear span of the union of $[\ell_{\rm nu}(\mathcal{P}_0)]$ as $\mathcal{P}_0$ ranges over $\mathcal{C}$. Write the tangent space $\dot{\mathcal{P}}_{\rm nu}$. Define $\ell^*_\beta = \ell_\beta - \Pi(\ell_\beta \mid \dot{\mathcal{P}}_{\rm nu})$ and ℓ^*_γ, likewise. These are called the *efficient score functions*. Writing $\ell^* = (\ell^{*\prime}_\beta, \ell^*_\gamma)'$, the information matrix in the semiparametric model is given by

$$I(P_0 \mid \beta, \gamma, \mathcal{P}) = E_{P_0}\, \ell^* \ell^{*\prime}.$$

In the discussion that follows we omit the subscript "0" in β_0, γ_0, σ_0^2, h_0 and g_0 which has been used to indicate they are the true values and functions. Also, we suppress the subscript "P_0" in E_{P_0}.

The following theorem exhibits ℓ^*_β and ℓ^*_γ for estimating β and γ. To state the theorem, let $c_t \equiv c_t(\gamma) = \sum_{j=0}^{t-1} \gamma^j$ and $\tilde{c} \equiv \tilde{c}(\gamma) = \sum_{t=1}^{r-1} c_t(\gamma)/r$. Write $X^w_{1t} \equiv X^w_{1t}(\gamma) = \sum_{j=0}^{t-1} \gamma^j X_{1,t-j}$ and $\tilde{X}^w_1 \equiv \tilde{X}^w_1(\gamma) = \sum_{t=1}^{r-1} X^w_{1t}(\gamma)/r$. Similarly, let $Z^w_{1t} \equiv Z^w_{1t}(\beta, \gamma) = \sum_{j=0}^{t-1} \gamma^j Z_{1,t-j}$ and $\tilde{Z}^w_1 \equiv \tilde{Z}^w_1(\beta, \gamma) = \sum_{t=1}^{r-1} Z^w_{1t}(\beta, \gamma)/r$. Define $\bar{X}_1 = \sum_{t=1}^{r} X_{1t}/r$.

Theorem 2.1 *The efficient score functions for estimating β and γ are given by*

$$\begin{aligned} \ell^*_\beta &= \sum_{t=1}^{r} (Z_{1t} - \bar{Z}_1) X_{1t}/\sigma^2 - \{w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)\}(\bar{X}_1 - E\bar{X}_1) \\ \ell^*_\gamma &= \sum_{t=1}^{r} (Z_{1t} - \bar{Z}_1) Y_{1,t-1}/\sigma^2 + \{\tilde{c}/(r-1)\sigma^2\} \sum_{t=1}^{r} (Z_{1t} - \bar{Z}_1)^2 \\ &\quad - \{w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)\}\{\beta'(\tilde{X}^w_1 - E\tilde{X}^w_1) + \tilde{Z}^w_1 - \tilde{c}\bar{Z}_1\} \end{aligned}$$

where $w^{(1)}$ denotes the first derivative of w.

The information matrix $I(P_0 \mid \beta, \gamma, \mathcal{P})$ can be calculated by using Theorem 2.1. Let $\Sigma_{\rm wtn} = \sum_{t=1}^{r} E(X_{1t} - \bar{X}_1)(X_{1t} - \bar{X}_1)'$ and $\Sigma_{\rm btn} = var(\bar{X}_1)$. Define $I_w = \int \{(w^{(1)}(z))^2/w(z)\}\, dz$. Then

$$E\, \ell^*_\beta \ell^{*\prime}_\beta = \sigma^{-2} \Sigma_{\rm wtn} + I_w \Sigma_{\rm btn}. \tag{2.6}$$

Define $\xi \equiv \xi(\gamma) = \sum_{t=1}^{r-1} \sum_{s=1}^{r-1} \sum_{j=1}^{t \wedge s - 1} \gamma^{|t-s| + 2j}$. It can be shown from a lengthy and cumbersome calculation that

$$E\, \ell^{*2}_\gamma \;=\; \beta' E\{\sum_{t=1}^{r-1} (X^w_{1t} - \tilde{X}^w_1)(X^w_{1t} - \tilde{X}^w_1)'\}\beta/\sigma^2 + 2\,\beta' \sum_{t=1}^{r-1} (c_t - \tilde{c}) E(X^w_{1t} - \tilde{X}^w_1) E(\bar{Z}_1)/\sigma^2$$

$$
\begin{aligned}
& + \sum_{t=1}^{r-1} (c_t - \tilde{c})^2 E(\bar{Z}_1^2)/\sigma^2 + I_w\{(\xi - r\tilde{c})\sigma^2/r^2 + \beta' var(\tilde{X}_1^w)\beta\} \qquad (2.7) \\
& + (1 - r^{-1}) \sum_{t=1}^{r-1} \sum_{j=0}^{t-1} \gamma^{2j} - \sum_{t=1}^{r-1} c_t^2/r - 2\,\tilde{c}^2/(r-1).
\end{aligned}
$$

$$
\begin{aligned}
E\,\ell_\beta^* \ell_\gamma^* \;=\; & \sigma^{-2}\{\sum_{t=1}^{r-1} \beta' E X_{1t}^w (X_{1,t+1} - \bar{X}_1) + \sum_{t=1}^{r-1} c_t E(X_{1,t+1} - \bar{X}_1) E(\bar{Z}_1)\} \qquad (2.8) \\
& + I_w \beta' E(\tilde{X}_1^w - E\tilde{X}_1^w)(\bar{X}_1 - E\bar{X}_1).
\end{aligned}
$$

The information matrix is readily obtained from (2.6), (2.7) and (2.8).

2.2 Construction of efficient estimators

Let $\theta = (\beta', \gamma)'$. Write $I = I(P_0 \,|\, \beta, \gamma, \mathcal{P})$. We construct an estimator $\hat{\theta}_n$ of θ such that $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in distribution to $\mathcal{N}(0, I^{-1})$. Define Z_{it} as we define Z_{1t} but replacing the subscript "1" by "i", i.e. $Z_{it} \equiv Z_{it}(\theta) = Y_{it} - \gamma Y_{i,t-1} - \beta' X_{it}$. Likewise, define $\bar{Z}_i$, X_{it}^w, $\tilde{X}_i^w$, Z_{it}^w and $\tilde{Z}_i^w$. Replace the subscript "1" by "i" in the formula for ℓ_β^* and ℓ_γ^* given at Theorem 2.1, and denote them by $\ell_{\beta,i}^*$ and $\ell_{\gamma,i}^*$, respectively. Define $\ell_i^* = (\ell_{\beta,i}^{*\prime}, \ell_{\gamma,i}^*)'$. Instead of writing just ℓ_i^* we will write $\ell_i^*(\theta)$ to stress its dependence on θ and for notational convenience in description of the efficient estimator given below. Note particularly that $\ell_i^*(\theta)$ depends on other parameters σ^2, h and g, too. Efficient estimators $\hat{\theta}_n$ are characterized by the following stochastic expansion:

$$
\hat{\theta}_n = \theta + n^{-1} I^{-1} \sum_{i=1}^{n} \ell_i^*(\theta) + o_p(n^{-1/2}). \qquad (2.9)
$$

We follow the usual one-step procedure for constructing an efficient estimator: (i) Find a $\sqrt{n}$-consistent estimator $\tilde{\theta}_n$ of θ. (ii) Assuming the true parameter value θ is known, find a reasonable estimator of σ^2, and using this construct an estimator of the density function $w(\cdot)$. (iii) Substitute the estimators obtained at (ii) into $\ell_i^*(\theta)$, and call it $\hat{\ell}_i^*(\theta)$. Also, construct an estimator of I using the estimators obtained at (ii), and denote it by $\hat{I}(\theta)$. (iv) Construct $\hat{\theta}_n$ by $\hat{\theta}_n = \tilde{\theta}_n + n^{-1} \hat{I}^{-1}(\tilde{\theta}_n) \sum_{i=1}^n \hat{\ell}_i^*(\tilde{\theta}_n)$.

First, we construct an initial estimator of θ which is $\sqrt{n}$-consistent. We take, as an initial estimator $\tilde{\theta}_n$, the minimizer of $\sum_{i=1}^n \sum_{t=1}^r \left\{(Y_{it} - \bar{Y}_i) - \gamma(Y_{i,t-1} - \bar{Y}_i) - \beta'(X_{it} - \bar{X}_i)\right\}^2$ with respect to β and γ where $\bar{X}_i = \sum_{t=1}^r X_{it}/r$ and $\bar{Y}_i = \sum_{t=1}^r Y_{it}/r$. Write $\upsilon_{it} = (X_{it}', Y_{i,t-1})'$, $m = \sum_{i=1}^n \sum_{t=1}^r \upsilon_{it} Y_{it}$ and $\mathcal{M} = \sum_{i=1}^n \sum_{t=1}^r \upsilon_{it} \upsilon_{it}'$. Then, the least squares initial estimator can be written as

$$
\tilde{\theta}_n = \mathcal{M}^{-1} m. \qquad (2.10)
$$

It can be shown that $\tilde{\theta}_n$ is $\sqrt{n}$-consistent.

Given the true value θ, we define $\tilde{\sigma}_n^2(\theta)$ by

$$\tilde{\sigma}_n^2(\theta) = \sum_{i=1}^{n}\sum_{t=1}^{r}\left\{(Y_{it} - \bar{Y}_i) - \gamma(Y_{i,t-1} - \bar{Y}_i) - \beta'(X_{it} - \bar{X}_i)\right\}^2 / n(r-1).$$

Next, we construct a density estimator $\hat{w}(\cdot;\theta)$. Recalling that w is the density of $\bar{Z}_i(\theta)$, we estimate it by a kernel estimator

$$\hat{w}(z;\theta) = n^{-1}\sum_{i=1}^{n} K_{b_n}(z - \bar{Z}_i(\theta)) + c_n$$

where $K_{b_n}(u) = (1/b_n)K(u/b_n)$, $K(u) = e^{-u}(1+e^{-u})^{-2}$ and b_n is a constant converging to zero at an appropriate rate to be described later. The constant c_n is introduced here to avoid technical difficulties due to zero denominators arising otherwise, and is taken to converge to zero too as n tends to infinity, whose rate is also to be specified below.

Now, define $\hat{\ell}_i^*(\theta) = (\hat{\ell}_{\beta,i}^{*\prime}(\theta), \hat{\ell}_{\gamma,i}^*(\theta))'$ where

$$\begin{aligned}
\hat{\ell}_{\beta,i}^*(\theta) &= \sum_{t=1}^{r}\left\{Z_{it}(\theta) - \bar{Z}_i(\theta)\right\} X_{it}/\tilde{\sigma}_n^2(\theta) \qquad (2.11)\\
&\quad - \left\{\hat{w}^{(1)}(\bar{Z}_i(\theta);\theta)/\hat{w}(\bar{Z}_i(\theta);\theta)\right\}(\bar{X}_i - \sum_{i=1}^{n}\bar{X}_i/n)\\
\hat{\ell}_{\gamma,i}^*(\theta) &= \sum_{t=1}^{r}\left\{Z_{it}(\theta) - \bar{Z}_i(\theta)\right\} Y_{1,t-1}/\tilde{\sigma}_n^2(\theta) + \left\{\tilde{c}(\gamma)/(r-1)\tilde{\sigma}_n^2(\theta)\right\}\sum_{t=1}^{r}\left\{Z_{it}(\theta) - \bar{Z}_i(\theta)\right\}^2\\
&\quad - \left\{\hat{w}^{(1)}(\bar{Z}_i(\theta);\theta)/\hat{w}(\bar{Z}_i(\theta);\theta)\right\}\Bigg\{\beta'\left(\tilde{X}_i^w(\gamma) - \sum_{i=1}^{n}\tilde{X}_i^w(\gamma)/n\right) \qquad (2.12)\\
&\qquad + \tilde{Z}_i^w(\theta) - \tilde{c}(\gamma)\bar{Z}_i(\theta)\Bigg\}.
\end{aligned}$$

One may estimate I by $n^{-1}\sum_{i=1}^n \hat{\ell}_i^*(\theta)\hat{\ell}_i^{*\prime}(\theta)$, or by substituting the unknown quantities, except θ, in the expressions given at (2.6), (2.7) and (2.8). It is well known that the latter approach yields more stable estimators, and so we proceed in that direction here. Denote by I_{11} the $d \times d$ left-top block of the information matrix I, and by I_{12} and I_{22}, the $d \times 1$ right-top and 1×1 right-bottom blocks, respectively. Let $\hat{\Sigma}_{\rm wtn} = n^{-1}\sum_{i=1}^n\sum_{t=1}^r(X_{it} - \bar{X}_i)(X_{it} - \bar{X}_i)'$ and $\hat{\Sigma}_{\rm btn} = n^{-1}\sum_{i=1}^n\{\bar{X}_i - \sum_{i=1}^n \bar{X}_i/n\}\{\bar{X}_i - \sum_{i=1}^n \bar{X}_i/n\}'$. Let $\hat{I}_w(\theta) = n^{-1}\sum_{i=1}^n\left\{\hat{w}^{(1)}(\bar{Z}_i(\theta);\theta)/\hat{w}(\bar{Z}_i(\theta);\theta)\right\}^2$. Define

$$\hat{I}_{11}(\theta) = \tilde{\sigma}^{-2}(\theta)\hat{\Sigma}_{\rm wtn} + \hat{I}_w(\theta)\hat{\Sigma}_{\rm btn}.$$

We estimate I_{12} by

$$\hat{I}_{12}(\theta) = \tilde{\sigma}^{-2}(\theta)\left\{\beta' n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{r-1} X_{it}^w(\gamma)(X_{i,t+1} - \bar{X}_i)\right.$$

$$+\left(n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{r-1}c_t(\gamma)(X_{i,t+1}-\bar{X}_i)\right)\left(n^{-1}\sum_{i=1}^{n}\bar{Z}_i(\theta)\right)\Bigg\}$$

$$+\beta'\hat{I}_w(\theta)n^{-1}\sum_{i=1}^{n}\left\{\tilde{X}_i^w(\gamma)-n^{-1}\sum_{i=1}^{n}\tilde{X}_i^w(\gamma)\right\}\left\{\bar{X}_i-n^{-1}\sum_{i=1}^{n}\bar{X}_i\right\}.$$

Finally, given the true value of θ, we construct an estimator of I_{22} by

$$\begin{aligned}\hat{I}_{22}(\theta) \;=\; & \tilde{\sigma}_n^{-2}(\theta)\beta'\left\{n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{r-1}\left(X_{it}^w(\gamma)-\tilde{X}_i^w(\gamma)\right)\left(X_{it}^w(\gamma)-\tilde{X}_i^w(\gamma)\right)'\right\}\beta \\ & +2\,\tilde{\sigma}_n^{-2}(\theta)\beta'\left\{n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{r-1}(c_t(\gamma)-\tilde{c}(\gamma))\left(X_{it}^w(\gamma)-\tilde{X}_i^w(\gamma)\right)\right\}\left\{n^{-1}\sum_{i=1}^{n}\bar{Z}_i(\theta)\right\} \\ & +\tilde{\sigma}_n^{-2}(\theta)\left\{n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{r-1}\left(c_t(\gamma)-\tilde{c}(\gamma)\right)^2\bar{Z}_i^2(\theta)\right\} \\ & +\hat{I}_w(\theta)\Bigg\{(\xi(\gamma)-r\tilde{c}(\gamma))\,\tilde{\sigma}_n^2/r^2 \\ & \quad +\beta' n^{-1}\sum_{i=1}^{n}\Big(\tilde{X}_i^w(\gamma)-n^{-1}\sum_{i=1}^{n}\tilde{X}_i^w(\gamma)\Big)\Big(\tilde{X}_i^w(\gamma)-n^{-1}\sum_{i=1}^{n}\tilde{X}_i^w(\gamma)\Big)'\beta\Bigg\} \\ & +(1-r^{-1})\sum_{t=1}^{r-1}\sum_{j=0}^{t-1}\gamma^{2j}-\sum_{t=1}^{r-1}c_t^2(\gamma)/r-2\,\tilde{c}^2(\gamma)/(r-1).\end{aligned}$$

Plugging the initial estimator $\tilde{\theta}_n$ into $\hat{\ell}_i^*(\theta)$ and $\hat{I}(\theta)$, we obtain the following estimator of θ:

$$\hat{\theta}_n=\tilde{\theta}_n+n^{-1}\hat{I}^{-1}(\tilde{\theta}_n)\sum_{i=1}^{n}\hat{\ell}_i^*(\tilde{\theta}_n) \tag{2.13}$$

The following theorem demonstrates that the estimator defined at (2.13) is a semiparametric efficient estimator of θ.

Theorem 2.2 *Assume that* $E(e^{t|\bar{X}_1|})<\infty$ *for some* $t>0$ *and that* $\int|u|^2h(u)\,du<\infty$. *If* $b_n\to 0, c_n\to 0$ *and* $nc_n^2b_n^6\to\infty$ *as* $n\to\infty$, *then*

$$\sqrt{n}(\hat{\theta}_n-\theta)\to\mathcal{N}(0,I^{-1})$$

in distribution as n tends to infinity.

3 Monte Carlo Simulations

The finite sample performances of the initial consistent [1] and the semiparametric efficient estimator are compared through the following Monte-Carlo (MC) scenarios.

We simulated samples of size $n = 20, 100, 1000$ with $r = 20, 50$ in a model with $d = 2$ regressors. In each MC sample, the regressors were generated according to a bivariate VAR model:

$$\begin{aligned} X_{it} &= RX_{i,t-1} + \eta_{it}, \ \text{ where } \eta_{it} \sim I\mathcal{N}_2(0, \sigma_X^2 I_2), \\ &\sigma_X = 1 \text{ and } R = \begin{pmatrix} 0.4 & 0.05 \\ 0.05 & 0.4 \end{pmatrix}. \end{aligned} \tag{2.14}$$

The simulation was initialized as follows: we chose $X_{i1} \sim \mathcal{N}_2(0, \sigma_X^2 \left(I_2 - R^2\right)^{-1})$ and start the iteration (2.14) for $t \geq 2$.

Then the obtained values of X_{it} were shifted around three different means to obtain almost 3 balanced groups of cross-sectional units from smaller to larger. We fixed $\mu_1 = (5,\ 5)'$, $\mu_2 = (7.5,\ 7.5)'$, $\mu_3 = (10,\ 10)'$. 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 $\gamma = 0.99,\ 0.90,\ 0.70,\ 0.10,$ 0.0, and $\sigma = 0.5$. For small values of γ 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 random effects α_i 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 $B = 1$. Although we do not pursue the interpretation of the effects in the empirical work below we have in the previous studies been interested in the use of such models to estimate firm specific efficiency levels. Since in such models the y is 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 $(1,\ 0.5)'$.

[1] In principal an efficient estimator designed along the lines of Ahn and Schmidt (1995) and Arellano and Bover (1995) could be used as our initial consistent estimator. We found, however, that the instability of such an estimator with the cross-section and time-series dimenions used in our Monte Carlo experiments rendered these initial consistent estimators too unrealiable. Instead we utilized a classical instrumental variables approach using lagged values of exogenous and endogenous variables and well as their first differences and lags in first differences in the spirit of Ahn and Schmidt, Arellano and Bover, and Anderson and Hsiao (1981, 1982).

Due to computing time limitations, most of the results were 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 the tables below, we display the sum of the two MC mean-squared errors for the $\beta' s$:

$$MSE = \sum_{j=1}^{2} \frac{1}{M} \sum_{m=1}^{M} (\theta_j^{0,m} - \theta_j)^2,$$

and the mean-square error for γ

$$MSE = \frac{1}{M} \sum_{m=1}^{M} (\theta_3^{0,m} - \theta_3)^2,$$

where θ_j^0 denotes either the initial estimator $\tilde{\theta}_j$ or the semiparametric efficient $\hat{\theta}_j$.

For the bandwidth b we selected an optimal fixed value b^* by running the whole Monte-Carlo experiment for a selected grid of 20 equally spaced values for b between 0.1 to 2.5. We report in the tables the results corresponding to the optimal bandwidth b^* which minimizes the MSE. In all the tried scenarios, the results were not very sensitive to the choice of b in the above grid. For the empirical analysis of the airline market data set in Section 4, we propose a data driven cross-validation algorithm.

n	**r**	$\widetilde{\gamma}$	$\widehat{\gamma}$	$\widetilde{\beta}$	$\widehat{\beta}$	$\mathbf{b}^*$
20	20	0.0032	0.0015	15.5502	8.8392	0.1
100	20	0.0006	0.0003	3.0454	1.8088	0.1
1000	20	0.0001	0.0000	0.2892	0.1687	0.1
20	50	0.0002	0.0001	7.5392	3.8261	0.1
100	50	0.0000	0.0000	1.6455	0.7319	0.1
1000	50	0.0000	0.0000	0.2387	0.0770	0.2

Table 1: *Monte Carlo MSE of the estimators of θ with M=500 replications. The figures for the MSE are multiplied by* 10^3. *Here* $\gamma = 0.99, \sigma = 0.5$, *and* $\mu_\alpha = 1$. *For n=1000 only M=100 replications were performed.*

n	**r**	$\widetilde{\gamma}$	$\widehat{\gamma}$	$\widetilde{\beta}$	$\widehat{\beta}$	$\mathbf{b}^*$
20	20	0.0228	0.0071	15.2719	8.9142	0.1
100	20	0.0044	0.0017	2.9824	1.8065	0.2
1000	20	0.0006	0.0002	0.2994	0.1699	0.1
20	50	0.0171	0.0029	8.6467	3.7017	0.1
100	50	0.0058	0.0006	2.0451	0.7543	0.1
1000	50	0.0030	0.0001	0.5500	0.0679	0.1

Table 2: *Monte Carlo MSE of the estimators of θ with M=500 replications. The figures for the MSE are multiplied by* 10^3. *Here* $\gamma = 0.90, \sigma = 0.5,$ *and* $\mu_\alpha = 1$. *For n=1000 only M=100 replications were performed.*

n	**r**	$\widetilde{\gamma}$	$\widehat{\gamma}$	$\widetilde{\beta}$	$\widehat{\beta}$	$\mathbf{b}^*$
20	20	1.0076	0.0980	20.8269	10.5991	0.2
100	20	0.4275	0.0194	5.8643	1.9396	0.1
1000	20	0.2959	0.0025	2.6676	0.1895	0.1
20	50	4.8821	0.2592	59.9342	4.4669	1.2
100	50	3.7212	0.0625	39.9281	0.9093	2.0
1000	50	3.4301	0.0300	36.7256	0.1089	1.4

Table 3: *Monte Carlo MSE of the estimators of θ with M=500 replications. The figures for the MSE are multiplied by* 10^3. *Here* $\gamma = 0.70, \sigma = 0.5,$ *and* $\mu_\alpha = 1$. *For n=1000 only M=100 replications were performed.*

n	**r**	$\widetilde{\gamma}$	$\widehat{\gamma}$	$\widetilde{\beta}$	$\widehat{\beta}$	$\mathbf{b}^*$
20	20	45.9048	7.6055	71.7682	15.6721	0.60
100	20	33.9708	1.6755	44.3691	3.4728	0.20
1000	20	32.3524	1.1484	39.8249	0.5260	1.60
20	50	177.2281	17.3109	251.9102	12.7766	0.40
100	50	166.2137	6.8044	225.0624	4.2812	1.30
1000	50	163.5490	4.7516	218.9318	2.4003	1.10

Table 4: *Monte Carlo MSE of the estimators of θ with M=500 replications. The figures for the MSE are multiplied by* 10^3. *Here* $\gamma = 0.10, \sigma = 0.5,$ *and* $\mu_\alpha = 1$. *For n=1000 only M=100 replications were performed.*

n	**r**	$\widetilde{\gamma}$	$\widehat{\gamma}$	$\widetilde{\beta}$	$\widehat{\beta}$	$\mathbf{b}^*$
20	20	58.4413	6.6230	76.3811	15.7628	0.50
100	20	43.0524	1.5662	46.4698	3.3897	0.20
1000	20	41.7777	0.8972	42.3493	0.4940	1.10
20	50	234.2394	17.0526	273.4654	13.1465	1.90
100	50	213.9966	5.5336	236.6754	3.9621	1.60
1000	50	210.3004	3.5021	229.8879	2.0487	2.00

Table 5: *Monte Carlo MSE of the estimators of θ with M=500 replications. The figures for the MSE are multiplied by* 10^3. *Here* $\gamma = 0.00, \sigma = 0.5,$ *and* $\mu_\alpha = 1$. *For n=1000 only M=100 replications were performed.*

As a global conclusion, it appears that our efficient estimator behaves well across the different MC scenarios even if γ is small. When autocorrelation is present our estimator increases the precision of the estimators of γ and β for the different sample sizes analyzed here.

4 Empirical Illustration

4.1 Data

In this section we illustrate our new estimator by estimating dynamic demand equations for a set of U. S. air carriers operating in a number of different markets (city-pairs) over time. The DB1A data on which our emprical illustration is based is a one in ten sample of all tickets issued from January 1979 through December 1992. These are aggregated so that all tickets with the same fare, airlines, plane changes and in the same quarter are grouped together. This leads to the loss of some potentially useful information such as date and time of day, flight number and equipment type. DOT further limits other information which appears on the actual ticket such as the complete fare class which includes restrictions on advance purchase, Saturday night stay, refundability and other restrictions. For this study, the only available fare information which is consistently given over time is whether the ticket is first class or coach and whether any restriction was placed on the ticket.

Before they eliminate information on time of day and date of travel, DOT uses this information to identify trip breaks, and consequently identifies the ultimate destination of travel. In some cases this can unambiguously be done: if there are two flight segments and the second ends where the first began, it is clearly a round trip ticket with the destination at the end of the first segment. If there are three trip segments things become less clear. The ultimate destination is most likely the place with the longest break. On the other hand, such a ticket could be associated with travel with both the city at the end of the first segment and the city at the end of the second segment were destinations. Over the time period of the study, DB1A included tickets with up to 23 segments. In some cases these are listed simply as round trip tickets. Since the researcher has less information than DOT had to make judgements about what the ultimate trip purpose was, he or she is in no position to "fix" these tickets. Still, leaving multi-destination tickets in the sample would seriously compromise the study, since they represent sales in more than one market. In this study only tickets that can be clearly classified as either one-way or round trips with a single destination are used. Up to six total segments are allowed for the flight (a five stop, one-way ticket or a round trip ticket with two stops per leg of the trip). Including only one way and round trip tickets with no more than six total segments eliminates just over 1% of the total data.

For clarity, it is important to define what a coupon segment is. If a consumer purchases a ticket, the itinerary for that ticket contains "coupons" that correspond to flight numbers. If the itinerary involves a change in flight numbers, then there is a new "coupon." Consequently a coupon segment is that part of the itinerary in which the flight number remains the same. It is important to note that this does not correspond perfectly with a stop along the trip. If there is a stop in an itinerary that leads to a change in flight number, then that stop leads to a change in coupon segments. This can occur when the passenger changes plains, but also can occur when the passenger stays on the plane. However, if the passenger stays on the plane during the stop and the airline does not change flight numbers, then there is no new coupon segment. This disconnect between coupon segments and stops will lead to some error when estimating marginal costs and when estimating the impact of the number of stops on price.

It is important to note that this study considers a market (route) to be neither the US as a whole nor, as in most studies, a trip between origin and destination airports. Instead, a market is considered to be a trip between origin and destination cities. Having the market defined as all flights in the US. leads one to the conclusion that regional carriers in different regions compete with each other. Since this is known not to be the case, it is prudent to conduct a more detailed analysis. However, defining a route by airports neglects the competition that airlines face from carriers that fly from different airports within the same city. For instance, if someone were to fly from Houston to Chicago, he or she would have four combinations of airport pairs to choose from: Bush Intercontinental/Chicago O'Hare, Bush Intercontinental/Chicago Midway, Houston Hobby/Chicago O'Hare, and finally Houston Hobby/Chicago Midway. If the flight were round trip, he or she would have four more airport combinations for the trip back. It is well known that United and American have a considerable share of flights from Bush Intercontinental to Chicago O'Hare because O'Hare is a hub for those two carriers. However, Southwest Airlines has a very large schedule of flights from Houston Hobby to Chicago Midway. Considering a route to be an airport pair would suggest that Southwest does not compete with United and American for flights between Houston and Chicago.

There are a number of factors for which controls other than standard demand variables such as own price, price of competitors, income, etc. are necessary in order to model the dynamic demand for airline travel. These are measured imperfectly. We imbed a number of these in the construction of the price index itself, following the methods outlined by Good et al. (2001). The factors can be categorized into five broad groups: Route specific effects, ticket restrictions, yield management, zero coupon tickets and network effects.

4.1.1 Route Specific factors

There are clearly other variables which many have attempted to incorporate into modeling the demand side of long distance travel. These include factors which are weather related, such as mean temperature difference, in an attempt to capture vacation travel in the winter months. Others have collected additional variables which attempt to capture the demand for business travel such as the number of white collar jobs in an area. We do have per capita income in the SMSA that surrounds one of the largest 80 airports as well as population and unemployment rate which we obtained from the BLS. We assume that other factors for which we have no controls are slow to change or that they are proxied well in the variables we do observed. The slowly moving factors that are markets (i.e., city–pair)-specific are captured with the random route effects which describe the origin-destination pair.

4.1.2 Ticket restrictions

A major feature of airline fare structures is ticket restrictions. These either increase the risk of travel for consumers (non-refundability) or provide the airlines with improved predictability about demand (advanced booking) and enhance their ability to provide price discrimination information by separating price sensitive consumers from business travelers with more inelastic demands (Saturday night stay-over). The major liability of using of DOT's DB1A as the primary source of ticket information is that it includes very incomplete information on ticket restrictions. There is typically a lag between fare type innovations and the way they are reported in DB1A. This makes it difficult to identify a consistent set of conditions under which service was accepted.

4.1.3 Yield management

There is a great deal of competition in published fares. It is not at all uncommon for different airlines providing service on the same route to offer similar fare classes (sets of fare restrictions) at an identical price. However, fare structures may not correspond to published fares, in part due to yield management practices. We attempt to capture the effect of yield management by controlling for the percentage of first class, first class restricted, and coach restricted tickets.

4.1.4 Zero coupon tickets

Frequent flyer miles were introduced in the mid 1980's. The practice has proven so successful that it has proliferated to other industries, even grocery stores offer discounts for frequent shoppers. To control for the effects of zero coupon tickets markups above marginal cost, the percentage of zero coupon tickets sold by the carrier for a particular route is controlled for in the construction of the price index.

4.1.5 Network Configuration

Much has been made out of changes in airline networks by increased use of hub-and-spoke type networks. Airlines find these network configurations useful because they allow for higher passenger densities on individual routes.

Indirect routing of passengers clearly benefits the airlines because they can provide travel to passengers with fewer flights, potentially taking advantage of economies of equipment size (larger aircraft tend to have lower costs per passenger mile) and higher load factors (filling otherwise empty seats on an aircraft cost the airline very little).

Many of the different network characteristics can be measured at the individual ticket level. The DB1A database allows identification of many of the characteristics of the trip. Most fundamentally, the origin of a trip can be identified as well as the ultimate destination as indicated by a trip break. Approximately 95% of trips are either one way or round trip (depending on the year) with a small number of multi-break tickets involving as many as 23 different flights. More complex routings tend to be slightly more prevalent in later years than in earlier ones. In order to gain an understanding of the bulk of trips, attention is limited to either one way or round trip tickets which are weighted by travel distance. Information from the DB1A also allows measurement of the number of segments in a ticket. To control for the effect of the number of segments in the itinerary, we also control for the percentage of tickets with any number of stops up to 5 stops.

The minimum number of segments for a one way ticket is one. By 1984, this number fell to 25%. A very different pattern emerges for round trip tickets which have a minimum of two segments. In 1979-1 the average number of segments was 2.8, this increased somewhat to 3.05 by 1992-4. At 3.0 it suggests that approximately half of the itineraries involved a change of planes on the outbound and inbound portions of the trip. The rationale behind the difference in the one way and round trip ticket patterns is not clear. It may suggest a correlation between one way and full fare tickets which have a higher quality of demanded service for the large premium in price. On the other hand, while the presumption behind round trip tickets is that they describe the full trip, that is not the case for one way tickets since the passenger will require, at the minimum, an additional ticket for the return flight. Consequently the presumption that a full fare ticket involves the ultimate destination seems less well founded.

4.2 Results

We utilize the following cross-validation method to select the bandwidth for our empirical study. Define $\widehat{w}_{-i}$ to be the density estimate constructed from all the $\bar{Z}_j$'s except $\bar{Z}_i$, that is to say,

$$\widehat{w}_{-i}(z) = \frac{1}{(n-1)b} \sum_{j \neq i} K\left(\frac{z - \bar{Z}_j}{b}\right).$$

Then, the log likelihood is averaged over each choice of omitted $\bar{Z}_i$ to give the score function

$$CV(b) = \frac{1}{n} \sum_{i=1}^{n} \log \widehat{w}_{-i}(\bar{Z}_i).$$

The likelihood cross-validation choice of b is then the value of b which maximizes the function $CV(b)$. Theoretical properties of this bandwidth selector was fully analyzed by Hall (1987).

We have analyzed the dynamic demands for upwards of 500 city-pair markets for 12 US carriers during the late 1970's through the early 1990's. These comprise the largest 80 markets in the US network. The airline carriers are: American, Continental, Delta, Northwest, Ozark, Piedmont, Republic, TWA, United, and US Air. Different carriers moved in and out of different markets during this period and we had to take this into account in selecting the periods and markets that would allow us to balance our panels for each carrier. Our estimator in principle could be modified to handle an unbalanced panel but we do not pursue that modification in this paper. The periods under study for the carriers are provided in Table 6.

Airline	**n**	**r**	**Obs.**	**Period**
American	43	63	2709	$79I - 94III$
Continental	16	57	912	$80III - 94III$
Delta	44	63	2772	$79I - 94III$
Eastern	59	18	1062	$83III - 87IV$
Frontier	22	16	352	$82I - 85IV$
Northwest	23	61	1403	$79I - 94I$
Ozark	20	28	560	$79I - 85IV$
Piedmont	25	30	750	$82I - 89II$
Republic	35	21	735	$81I - 86I$
TWA	40	57	2280	$80III - 94III$
United	48	59	2832	$80I - 94III$
USAir	34	63	2142	$79I - 94III$

Table 6: *Summary of airlines, number of markets (n), time periods (r) and time intervals for the sample*

Summary statistics for each carrier are in Table 7. Demand is measured in millions of revenue ton miles, the price is measured in terms of price/revenue ton mile, and population is in thousands.

Airline	ln(**demand**)	ln(**price**)	ln(**population**)
American	18.29(1.003)	−2.723(0.199)	7.535(0.951)
Continental	18.22(1.108)	−2.977(0.227)	7.603(0.669)
Delta	18.24(0.839)	−2.587(0.256)	7.432(0.968)
Eastern	17.57(1.070)	−2.632(0.338)	7.286(0.916)
Frontier	16.63(0.701)	−2.710(0.179)	6.889(0.843)
Northwest	17.91(1.156)	−2.895(0.237)	8.010(0.930)
Ozark	16.31(0.952)	−2.527(0.183)	7.496(0.844)
Piedmont	17.32(0.773)	−2.510(0.232)	7.619(0.907)
Republic	16.64(1.081)	−2.459(0.273)	7.410(0.824)
TWA	17.57(0.946)	−2.768(0.288)	7.591(0.913)
United	18.19(1.225)	−2.816(0.266)	7.454(0.916)
USAir	17.53(0.962)	−2.490(0.249)	7.591(0.918)

Table 7: *Means and standard deviations(in parentheses) for the variables in the demand equations by airline.*

Results for the the semi-parametric efficient estimator and the IV-estimator estimator are in Table 8. Our specification is rather parsimonous with ln(*demand*), a function of ln(*demand*)$_{-1}$, ln(*own price*), and ln(*population*) in the originating city. City-pair characteristics as well as those portions of demand characteristics for which we have no explicit controls are modeled as random effects. Our results suggest that most city-pair markets have inelastic short-run of demand and quite elastic long-run demands. Competition from other carriers in terms of significant cross elastity of demand cannot be identified. Our results are reasonable with no evidence that the roots of the dynamic equation are unstable. Demand appears to adjust reasonably quickly to price changes but there is scope for significant market power to be exercised in the short-run.

Airline	ln(**demand**)$_{-1}$	ln(**price**)	ln(**price**)$_{comp}$	ln(**population**)
American − spe	0.9146(0.0011)	−0.4209(0.0345)	−0.0345(0.0331)	−0.0669(0.0081)
American − iv	0.9156(0.0071)	−0.3524(0.0339)	0.0204(0.0405)	0.0367(0.0525)
Contintental − spe	0.6830(0.0091)	−2.0407(0.1422)	0.2299(0.1575)	0.1650(0.0644)
Continental − iv	0.7182(0.0259)	−1.3268(0.1480)	0.1633(0.1876)	0.1736(0.1244)
Delta − spe	0.8822(0.0009)	−0.3686(0.0308)	−0.1096(0.0275)	0.1177(0.0080)
Delta − iv	0.8855(0.0086)	−0.3483(0.0301)	−0.0326(0.0301)	0.0045(0.0399)
Northwest − spe	0.9231(0.0027)	−0.2472(0.0733)	−0.2809(0.0829)	0.1291(0.0220)
Northwest − iv	0.9297(0.0124)	−0.1812(0.0764)	−0.0832(0.0998)	0.0200(0.0650)
Ozark − spe	0.7551(0.0041)	0.1414(0.1221)	−0.0797(0.1383)	0.0878(0.0337)
Ozark − iv	0.7646(0.0210)	−0.6255(0.1276)	0.3016(0.1483)	0.0032(0.0801)
Piedmont − spe	0.9210(0.0019)	−0.2361(0.0447)	−0.0756(0.0638)	0.0016(0.0158)
Piedmont − iv	0.9219(0.0168)	−0.1907(0.0514)	−0.0756(0.0638)	0.0071(0.0198)
Republic − spe	0.8053(0.0037)	−0.3271(0.1242)	−0.2147(0.1476)	0.0773(0.0368)
Republic − iv	0.8133(0.0254)	−0.5481(0.1239)	0.1571(0.1660)	0.0807(0.0563)
TWA − spe	0.8553(0.0019)	−0.2696(0.0456)	−0.0896(0.0556)	0.0347(0.0140)
TWA − iv	0.8578(0.0122)	−0.2232(0.0467)	−0.0654(0.0676)	0.0484(0.0313)
United − spe	0.9095(0.0014)	−0.4475(0.0363)	−0.0566(0.0360)	0.0470(0.0110)
United − iv	0.9124(0.0079)	−0.3418(0.0369)	0.0212(0.0503)	0.0311(0.0254)
USAir − spe	0.9168(0.0013)	−0.3937(0.0404)	−0.1912(0.0492)	0.0099(0.0143)
USAir − iv	0.9197(0.0092)	−0.3303(0.0410)	−0.1274(0.0520)	−0.0049(0.0557)

Table 8: *Parameter Estimates and standard deviations for the dynamic demand equations by airline.*

5 Concluding Remarks

In this paper we have introduced a new class of estimator for the dynamic panel data models Our semiparametric efficient estimator appears to perform well in finite sample Monte Carlo comparisons with competing estimators. We illustrate its use in an analysis of dynamic demands for airline service in selected city-pair markets in the US domestic industry. We find evidence that firms are operating on the inelastic portion of the demand schedule, a result consistent with significant short-run market power or collusive behavior. We also find substantial differences between short-run and long-run estimated demands by consumers in these markets, suggesting that such dynamic adjustments should be taken into consideration in analyzing competition policy and market behavior models in this important and highly litigious industry.

References

[1] Ahn, S. and P. Schmidt (1995), "Efficient estimation of a model with dynamic panel data", *Journal of Econometrics*, 68, 5–27.

[2] Alam, I. Semenick, and R. C. Sickles (2000), "A time series analysis of deregulatiory dynamics and technical efficiency: the case of the U. S. airline industry", *International Economic Review*, 41, 203–218.

[3] Anderson, T. W. and C. Hsiao (1981), "Estimation of dynamic models with error components", *Journal of the American Statistical Association*, 76, 598–606.

[4] Anderson, T. W. and C. Hsiao (1982), "Formulation and estimation of dynamic models using panel data", *Journal of Econometrics*, 18 (1982): 47–82, reprinted in *The Econometrics of Panel Data*, ed. by G.S. Maddala (1993), Brookfield VT: Edward Elgar.

[5] Arellano, M. and S. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *The Review of Economic Studies*, 58, 277–297.

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

[7] Baltagi. B. (1995), *The Econometric Analysis of Panel Data.* New York: John Wiley and Sons.

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

[9] Good, D., R. C. Sickles, and J. C. Weiher (2001), "On a new hedonic price index for airline travel", mimeo, Rice University.

[10] Hall, P. (1987), "On Kullback-Leibler loss and density estimation", *The Annals of Statistics*, 15, 1491–1519.

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

[12] L. Mátyás, and P. Sevestre (1996), *The Econometrics of Panel Data: A Handbook of the Theory with Applications.* Boston: Kluwer.

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

[14] Park, B. U., R. C. Sickles, and L. Simar (1998), "Stochastic frontiers: A semiparametric approach", *Journal of Econometrics*, 84, 273–301.

[15] Park, B. U., R. C. Sickles, and L. Simar (2003), "Semiparametric efficient estimation of AR(1) panel data models", forthcoming in the *Journal of Econometrics*.

Appendix

A.1 Proof of Theorem 2.1.

The score functions are given by

$$\begin{aligned}
\ell_\beta &= \sum_{t=1}^{r}(Z_{1t}-\bar Z_1)X_{1t}/\sigma^2 - \{w^{(1)}(\bar Z_1)/w(\bar Z_1)\}\bar X_1,\\
\ell_\gamma &= \sum_{t=2}^{r}(Z_{1t}-\bar Z_1)Y_{1,t-1}/\sigma^2 - \{w^{(1)}(\bar Z_1)/w(\bar Z_1)\}\tilde Y_1,\\
\ell_{\sigma^2} &= (2\sigma^2)^{-1}\left\{\sum_{t=1}^{r}(Z_{1t}-\bar Z_1)^2/\sigma^2\right.\\
&\quad \left. + \int \left(\bar\sigma^{-2}(\bar Z_1-u)^2 - r\right)\bar\sigma^{-1}\phi\left((\bar Z_1-u)/\bar\sigma\right) h(u)\,du/w(\bar Z_1)\right\},
\end{aligned}$$

where $\tilde Y_1 = \sum_{t=1}^{r-1} Y_{1t}/r$ and $\phi(\cdot)$ is the density function of the standard normal distribution. The tangent space $\dot{\mathcal{P}}_{\rm nu}$ may be decomposed into V_1, V_2 and V_3, i.e. $\dot{\mathcal{P}}_{\rm nu} = V_1+V_2+V_3$, where $V_1 = [\ell_{\sigma^2}]$ and

$$V_2 = \{a(\bar Z_1)\in L_2(P_0)\,:\, Ea(\bar Z_1)=0\},\ V_3 = \{b(X_1)\in L_2(P_0)\,:\, Eb(X_1)=0\}.$$

The following lemma shows that ℓ_β and ℓ_γ are perpendicular to V_3.

Lemma A.1 $E(\ell_\beta\,|\,X_1)=0$ *and* $E(\ell_\gamma\,|\,X_1)=0$.

Proof. Note that $\{Z_{1t}-\bar Z_1\}$, $\bar Z_1$ and X_1 are independent. Since $E(Z_{1t}-\bar Z_1)=0$ and $E\{w^{(1)}(\bar Z_1)/w(\bar Z_1)\} = \int w^{(1)}(u)\,du = 0$, we obtain $E(\ell_\beta\,|\,X_1)=0$. Next, we prove the second part. We can write $Y_{1t} = \beta' X_{1t}^w + c_t\alpha_1 + \sum_{j=0}^{t-1}\gamma^j\varepsilon_{1,t-j}$. Thus

$$E(\ell_\gamma\,|\,X_1) = \sigma^{-2}\sum_{t=1}^{r-1}\sum_{j=0}^{t-1}\gamma^j E\{\varepsilon_{1,t-j}(\varepsilon_{1,t+1}-\bar\varepsilon_1)\} - E\{\tilde Y_1 w^{(1)}(\bar Z_1)/w(\bar Z_1)\,|\,X_1\}. \tag{A.1}$$

The first term in (A.1) equals $\sigma^{-2}\sum_{t=1}^{r-1}\sum_{j=0}^{t-1}\gamma^j(-\sigma^2/r) = -\tilde c$. For the second term, note $\tilde Y_1 = \beta'\tilde X_1^w + \tilde c\bar Z_1 + r^{-1}\sum_{j=0}^{t-1}\gamma^j(Z_{1,t-j}-\bar Z_1)$. By this and the facts that $E\{w^{(1)}(\bar Z_1)/w(\bar Z_1)\}=0$ and that $E(Z_{1,t-j}-\bar Z_1)=0$, the second term equals

$$\tilde c E\{\bar Z_1 w^{(1)}(\bar Z_1)/w(\bar Z_1)\} = \tilde c\int u w^{(1)}(u)\,du = -\tilde c. \quad \text{(q.e.d.)}$$

Lemma A.1 implies that writing $W = [\ell_{\sigma^2} - \Pi(\ell_{\sigma^2}\,|\,V_2)]$

$$\begin{aligned}
\ell_\beta^* &= \ell_\beta - \Pi(\ell_\beta\,|\,V_2) - \Pi\{\ell_\beta - \Pi(\ell_\beta\,|\,V_2)|W\}\\
\ell_\gamma^* &= \ell_\gamma - \Pi(\ell_\gamma\,|\,V_2) - \Pi\{\ell_\gamma - \Pi(\ell_\gamma\,|\,V_2)|W\}.
\end{aligned}$$

We compute ℓ_β^* first. Note that $\Pi(\ell_\beta \,|\, V_2) = E(\ell_\beta \,|\, V_2) = -E(\bar{X}_1)w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)$. Thus

$$\ell_\beta - \Pi(\ell_\beta \,|\, V_2) = \sigma^{-2}\sum_{t=1}^{r}(Z_{1t} - \bar{Z}_1)X_{1t} - \{w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)\}(\bar{X}_1 - E\bar{X}_1). \tag{A.2}$$

Since $E\sum_{t=1}^{r}(Z_{1t} - \bar{Z}_1)^2 = (r-1)\sigma^2$, we obtain

$$\ell_{\sigma^2} - \Pi(\ell_{\sigma^2} \,|\, V_2) = (2\sigma^4)^{-1}\{\sum_{t=1}^{r}(Z_{1t} - \bar{Z}_1)^2 - (r-1)\sigma^2\}. \tag{A.3}$$

Now by symmetry of the distribution of $(Z_{1t} - \bar{Z}_1)$ and by independence of $Z_{1t} - \bar{Z}_1$, $\bar{Z}_1$ and X_1, it follows that $E\{\sum_{t=1}^{r}(Z_{1t} - \bar{Z}_1)X_{1t}\}\{\sum_{t=1}^{r}(Z_{1t} - \bar{Z}_1)^2\} = 0$ and

$$E\{w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)\}(\bar{X}_1 - E\bar{X}_1)\}\{\sum_{t=1}^{r}(Z_{1t} - \bar{Z}_1)^2 - (r-1)\sigma^2\} = 0.$$

Thus, $\ell_\beta - \Pi(\ell_\beta \,|\, V_2)$ is perpendicular to $\ell_{\sigma^2} - \Pi(\ell_{\sigma^2} \,|\, V_2)$, which implies that $\ell_\beta^* = \ell_\beta - \Pi(\ell_\beta \,|\, V_2)$. The formula for ℓ_β^* follows from (A.2).

Next, we compute ℓ_γ^*. By independence of $Z_{1t} - \bar{Z}_1$, $\bar{Z}_1$ and X_1, we have

$$\begin{aligned} E(\ell_\gamma \,|\, V_2) &= \sigma^{-2}\sum_{t=1}^{r-1}\sum_{j=0}^{t-1}\gamma^j E(\varepsilon_{1,t-j} - \bar{\varepsilon}_1)(\varepsilon_{1,t+1} - \bar{\varepsilon}_1) - \{w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)\}\left\{\beta' E(\tilde{X}_1^w) + \tilde{c}\bar{Z}_1\right\} \\ &= -\tilde{c} - \{w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)\}\left\{\beta' E(\tilde{X}_1^w) + \tilde{c}\bar{Z}_1\right\}. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} \ell_\gamma - \Pi(\ell_\gamma \,|\, V_2) &= \sigma^{-2}\sum_{t=2}^{r}(Z_{1t} - \bar{Z}_1)Y_{1,t-1} + \tilde{c} \\ &\quad -\{w^{(1)}(\bar{Z}_1)/w(\bar{Z}_1)\}\Big\{\beta'(\tilde{X}_1^w - E\tilde{X}_1^w) + \sum_{t=1}^{r-1}\sum_{j=0}^{t-1}\gamma^j(Z_{1,t-j} - \bar{Z}_1)/r\Big\}. \end{aligned}$$

To calculate $\Pi\left(\ell_\gamma - \Pi(\ell_\gamma \,|\, V_2) \,|\, W\right)$, we find

$$E\{\ell_\gamma - \Pi(\ell_\gamma \,|\, V_2)\}\{\ell_{\sigma^2} - \Pi(\ell_{\sigma^2} \,|\, V_2)\} = -\tilde{c}/\sigma^2 \tag{A.4}$$

$$E\{\ell_{\sigma^2} - \Pi(\ell_{\sigma^2} \,|\, V_2)\}^2 = (r-1)/(2\sigma^4). \tag{A.5}$$

Denote the left hand sides of (A.4) and (A.5) by ζ_{12} and ζ_{22}, respectively. Then from (A.3), (A.4) and (A.5), we obtain

$$\begin{aligned} \Pi\left(\ell_\gamma - \Pi(\ell_\gamma \,|\, V_2) \,|\, W\right) &= (\zeta_{12}/\zeta_{22})\{\ell_{\sigma^2} - \Pi(\ell_{\sigma^2} \,|\, V_2)\} \\ &= -\{\tilde{c}/(r-1)\sigma^2\}\sum_{t=1}^{r}(Z_{1t} - \bar{Z}_1)^2 + \tilde{c}, \end{aligned}$$

which leads to the formula for ℓ_γ^*.

A.2 Proof of Theorem 2.2

Define

$$w_n(z) \equiv w_n(z;\sigma^2) \equiv w_n(z;\sigma^2,h) = K_{b_n} * w(z;\sigma^2,h) + c_n$$

where * denotes the convolution. We write $\hat{r}$, r_n and r for $\hat{w}^{(1)}/\hat{w}$, $w_n^{(1)}/w_n$ and $w^{(1)}/w$, respectively. Define $I_{w,n} = \int r_n^2(z)w(z)\,dz$. Define I_n as in the definition of the information matrix I but with I_w being replaced by $I_{w,n}$. Following the arguments for the proof of (B.9) in Park, Sickles and Simar (1998), one can verify

$$E\left\{r_n(\bar{Z}_1(\theta)) - r(\bar{Z}_1(\theta))\right\}^2 \to 0. \tag{A.6}$$

It follows from (A.6) that $I_n \to I$ as n tends to infinity.

Now, it may be proved that

$$\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i - E\bar{X}_1\right) n^{-1/2}\sum_{i=1}^{n} r_n(\bar{Z}_i(\theta)) \to 0 \tag{A.7}$$

$$n^{-1/2}\sum_{i=1}^{n}(\bar{X}_i - E\bar{X}_1)\{r_n(\bar{Z}_i(\theta)) - r(\bar{Z}_i(\theta))\} \to 0, \tag{A.8}$$

both in the sense of convergence in probability. They follows since the left hand sides of (A.7) and (A.8) have zero means by independence of $\bar{X}_i$ and $\bar{Z}_i(\theta)$, and variances bounded by $n^{-1}var(\bar{X}_1)I_{w,n}$ and $var(\bar{X}_1)E\{r_n(\bar{Z}_1(\theta)) - r(\bar{Z}_1(\theta))\}^2$, respectively, both of which converge to zero as n tends to infinity. Similarly, it can be shown that

$$\left(n^{-1}\sum_{i=1}^{n}\tilde{X}_i^w(\gamma) - E\tilde{X}_1^w(\gamma)\right) n^{-1/2}\sum_{i=1}^{n} r_n(\bar{Z}_i(\theta)) \to 0 \tag{A.9}$$

$$n^{-1/2}\sum_{i=1}^{n}\left(\tilde{X}_i^w(\gamma) - E\tilde{X}_1^w(\gamma)\right)\{r_n(\bar{Z}_i(\theta)) - r(\bar{Z}_i(\theta))\} \to 0, \tag{A.10}$$

both in the sense of convergence in probability.

Define $\check{\ell}^*_{\beta,i}(\theta)$ and $\check{\ell}^*_{\gamma,i}(\theta)$ as in the definitions of $\hat{\ell}^*_{\beta,i}(\theta)$ and $\hat{\ell}^*_{\beta,i}(\theta)$ at (2.11) and (2.12), respectively, with $\hat{w}(\bar{Z}_i(\theta);\theta)$ being replaced by $w_n(\bar{Z}_i(\theta);\sigma^2)$ and $\tilde{\sigma}_n^2$ by σ^2, and let $\check{\ell}^*_i(\theta) = (\check{\ell}^{*\prime}_{\beta,i}(\theta), \check{\ell}^*_{\gamma,i}(\theta))'$. Then, (A.7) $\sim$ (A.10) imply

$$n^{-1/2}I_n^{-1}\sum_{i=1}^{n}\check{\ell}^*_i(\theta) \to \mathcal{N}(0, I^{-1}) \tag{A.11}$$

in distribution as n tends to infinity. Now, it can be shown that as in the proofs of Lemma A.2 and (A.16) of Park and Simar (1994)

$$\hat{I}(\tilde{\theta}_n) - I_n \to 0, \tag{A.12}$$

$$n^{-1/2}\sum_{i=1}^{n}\left(\bar{X}_i - n^{-1}\sum_{i=1}^{n}\bar{X}_i\right)\left\{\hat{r}(\bar{Z}_i(\tilde{\theta}_n);\tilde{\theta}_n) - r_n(\bar{Z}_i(\tilde{\theta}_n);\sigma^2)\right\} \to 0, \tag{A.13}$$

$$n^{-1/2}\sum_{i=1}^{n}\left(\tilde{X}_i^w(\tilde{\gamma}_n) - n^{-1}\sum_{i=1}^{n}\tilde{X}_i^w(\tilde{\gamma}_n)\right)\left\{\hat{r}(\bar{Z}_i(\tilde{\theta}_n);\tilde{\theta}_n) - r_n(\bar{Z}_i(\tilde{\theta}_n);\sigma^2)\right\} \to 0, \tag{A.14}$$

$$n^{-1/2}\sum_{i=1}^{n}\left(\tilde{Z}_i^w(\tilde{\theta}_n) - \tilde{c}(\tilde{\gamma}_n)\bar{Z}_i(\tilde{\theta}_n)\right)\left\{\hat{r}(\bar{Z}_i(\tilde{\theta}_n);\tilde{\theta}_n) - r_n(\bar{Z}_i(\tilde{\theta}_n);\sigma^2)\right\} \to 0, \tag{A.15}$$

all in the sense of convergence in probability. Since $\tilde{\sigma}_n^2(\tilde{\theta}_n)$ converges to σ^2 in probability, (A.13) $\sim$ (A.15) imply

$$n^{-1/2}I_n^{-1}\sum_{i=1}^{n}\left\{\hat{\ell}_i^*(\tilde{\theta}_n) - \check{\ell}_i^*(\tilde{\theta}_n)\right\} \to 0 \tag{A.16}$$

in probability. The theorem follows then from (A.11), (A.12) and (A.16) since for any $C > 0$

$$\sup\left\{\left|n^{-1/2}\sum_{i=1}^{n}\{\check{\ell}_i^*(\theta') - \check{\ell}_i^*(\theta) + I_n(\theta' - \theta)\right| \,:\, n^{1/2}|\theta' - \theta| \le C\right\} \to 0$$

in probability which can be proved as in Park and Simar (1994).