Title
Essays on Nonparametric Estimation of Dynamic Models

Author
Kang, David Minkee

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Nonparametric Estimation of Dynamic Models

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy in Economics

by

David Minkee Kang

2012

ABSTRACT OF THE DISSERTATION

Essays on Nonparametric Estimation of Dynamic Models

by

David Minkee Kang

Doctor of Philosophy in Economics

University of California, Los Angeles, 2012

Professor Rosa L. Matzkin, Chair

In this dissertation we describe conditions for nonparametric identification and methods for estimating dynamic simultaneous equation models. These models have two distinct sources of endogeneity: lagged dependent variables that are related to autocorrelated unobservable variables and endogeneity through a simultaneous equations structure. Until now, nonparametric estimation has been limited to models with either one or the other. In the first chapter we show that the structural functions in such models are identified with panel data under assumptions commonly made in nonparametric econometrics. We do so by borrowing intuition from existing literature on dynamic panel models. In the second chapter of the dissertation we describe conditions needed for consistent and asymptotically normal nonparametric estimation of dynamic simultaneous equations models. In the third chapter we nonparametrically estimate dynamic demand functions for airline travel using recent data.

While previous work relies on the linearity of the demand function (or other parametric assumptions), the functions we estimate are fully nonparametric. The nonparametric estimates exhibit promising out-of-sample forecast properties when compared to linear models in a limited forecasting exercise.

The dissertation of David Minkee Kang is approved.

Jinyong Hahn

Roger Farmer

Connan Snider

Sarah Reber

Rosa L. Matzkin, Committee Chair

University of California, Los Angeles

2012

The author would like to dedicate this dissertation to his family.

Contents

David Minkee Kang attended the Massachusetts Institute of Technology, where he received an SB in Course XIV (Economics). He then worked at the Federal Reserve Bank of Chicago with Drs. John Fernald, Francois Velde and Jeff Campbell on economic research and participated in the FOMC process. He holds a Master of Science degree from the University of California, Los Angeles in Economics.

Nonparametric Identification of Dynamic Simultaneous Equations Models

1 Introduction

In this chapter we describe nonparametric identification and estimation in dynamic simultaneous equation models in which the structural functions are nonseparable in unobservable variables. We are particularly interested in environments where the unobservable variables are not independent over time - they are autocorrelated. These models are not straightforward to estimate because they have two distinct sources of endogeneity: simultaneous equations structure and lagged dependent variable regressors (which are correlated with the autocorrelated unobservable variables). Until now, nonparametric estimation has been limited to models with either one or the other. The contribution of this chapter is adapting linear panel model intuition to show that identification in dynamic nonparametric simultaneous models is achieved under assumptions that are commonly made in nonparametric econometrics. One advantage of this approach to identification is that we can apply existing estimation techniques with known asymptotic properties.

There are a number of reasons that these particular models are important in economics. First, there are many important economic relationships that are thought to be determined through simultaneous equations. Some have called the simultaneous equation model "perhaps the most remarkable development in econometrics" (cf. Hausman (1983)). Additionally, economic models are often characterized with dynamics - choices or conditions in prior periods are important for choices in the future. Dynamic models using panel data are also necessary for separating the effects of state dependence from unobserved heterogeneity. Third, we focus on nonparametric identification (and estimation) because there are environments in which we think unobservable variables have nontrivial influence on marginal effects. We want to ensure that estimation in those cases is robust to model misspecification.

In this chapter we use theory and intuition from research on (static) simultaneous equations models, dynamic models and nonparametric models. This chapter might be considered a nonparametric analog to Bhargava and Sargan's (1983) paper where they propose treating dynamic panel models as a system of linear simultaneous equations model. In the same way we extend the nonparametric simultaneous equation identification results of Matzkin (2008) and the corresponding estimation results in Matzkin (2010) to dynamic models with the use of conditional independence conditions and by treating the set of simultaneous equations in each period as an element in a larger simultaneous equations structure.

For readers that are familiar with panel models, we are focusing on environments with observations on panel data of N observations over T time periods, where T is small relative to N. While we use panel data to identify the structural elements of the model, we are not dealing with what would be considered a 'standard' dynamic panel model. In this chapter, for instance, we assume that autocorrelated unobservable variables are related to the lagged dependent variables but not to some of the other included regressors. In companion papers we extend the results of this paper to include functional forms and distributional assumptions that are more typical of the dynamic panel literature.

There have been several recent papers on identifying nonparametric dynamic models using panel data. Cunha, Heckman and Schennach (2010) showed identification of a nonparametric dynamic structural function where the outcome variables were unobservable measures of 'ability.' This chapter is also about the identification of nonparametric laws of motion where lagged dependent variables are correlated with autocorrelated unobservables, but it is in an environment where the dependent variables are determined by simultaneous equations. Also, the focus of this paper is not on the evolution of unobservable variables, though we do describe conditions under which functions determining the evolution of unobservable variables are identified from data. Hu and Shum (2010) and Shiu and Hu (2010) also showed identification in dynamic processes with autocorrelated unobservable variables but for models with scalar outcomes and without identification of structural features of the model as in this chapter.

This chapter draws on existing work on identification in nonparametric models with endogenous regressors. The dynamics in the models we examine roughly translate to a system of triangular simultaneous equations. Chesher (2003) identifies the derivatives of the structural function using a control function approach that uses local conditional independence conditions.[1] Imbens and Newey (2009) use global conditional independence assumptions and support conditions in a similar model to identify and estimate 'structural effects.' They

[1] The actual assumption of 'quantile insensitivity' in Chesher (2003) is weaker. For further discussion and a survey on endogeneity in nonparametric models, please see Blundell and Powell (2006).

discuss the necessity of large variation in instruments for identification; our results also rely on large variation in our X_t variables.

This chapter also contributes to the literature on nonparametric identification using panel data. Arellano and Honore (2001), in writing about estimation of linear models with lagged dependent variables, noted that "almost nothing is known" about nonlinear panel models with non-exogenous variables. Since then, many authors have relaxed the assumption of regressor exogeneity in such models. For instance, Altonji and Matzkin (2005) identified features of nonparametric panel models with conditional density restrictions. Lee (2010) outlines identification and estimation in a nonparametric panel fixed effects model with additive unobservable terms. Evdokimov (2009) estimated a model where a random effect is nonseparable from X, but with an additive error term. Graham and Powell (2010) showed identification of "average partial effects" in panel models with nonseparable individual heterogeneity. Bester and Hansen (2007) showed identification in semiparametric correlated random effects models where individual effects are a nonparametric. These papers differ from this chapter in that we do not assume linearity in the outcome equations and we focus on endogeneity due to lagged dependent variables.

The model is described in greater detail in the following section, along with our distributional and functional assumptions. Also in Section 2, we state the main theorem and its proof. Section 3 contains our estimation results, and Section 4 concludes.

2 Model

We consider models of the form

$$Y_{it}^1 = m_1(Y_{it}^2, Y_{i,t-1}^1, Y_{i,t-1}^2, X_{it}^1 + \varepsilon_{it}^1)$$

$$Y_{it}^2 = m_2(Y_{it}^1, Y_{i,t-1}^1, Y_{i,t-1}^2, X_{it}^2 + \varepsilon_{it}^2)$$

The models that we focus on are dynamic; lagged dependent variables directly affect current values. Standard nonparametric methods would be valid except that the dependent variables are determined according to a simultaneous equation system as well as the fact that the unobserved error terms are correlated over time. In order to identify the structural equations in this model we utilize panel data: N observations indexed by $i = 1, \dots N$, for T periods of panel data, indexed by $t = 0, \dots T$. The discussions in this chapter will be restricted to 'short panel' situations where T is small relative to N.

We consider unobservable variables that evolve according to a simple nonseparable autoregressive process:

$$\varepsilon_{it}^1 = h_1(\varepsilon_{i,t-1}^1, \eta_{it}^1)$$

$$\varepsilon_{it}^2 = h_2(\varepsilon_{i,t-1}^2, \eta_{it}^2),$$

The overall structure makes the regressor $Y_{i,t-1}^1$ correlated with ε_{it}^1, since $\varepsilon_{i,t-1}^1$ enters the equation for $Y_{i,t-1}^1$ directly.

An important feature of these models is that unobserved heterogeneity is not restricted to be additive. In particular, the marginal effect of y_{t-1} on y_t is allowed to vary with ε_t structurally; that is, the effect of y_{t-1} is separately identified from the effect of ε_{t-1}'s influence on ε_t, but this effect is allowed to depend on ε_t. In a model of 'education production,' for instance, this means that the effect of prior academic success on current achievement depends on a child's ability but attribution of this effect is not due to the endogeneity problem (since prior academic success will be associated with higher ability). In addition, the unobserved variables evolve over time according to a nonparametric function that we can also identify. We do not discuss explicitly issues of stability so we refrain from making comparisons between the form of unobserved heterogeneity in this paper and the standard notions of random or fixed effects.[2]

To simplify notation, we will refer to $Y_{it} = \begin{bmatrix} Y_{it}^1 & Y_{it}^2 \end{bmatrix}'$ (same for X_{it}, ε_{it}, η_{it}). We adopt the same shorthand for functions ($m(\cdot)$ will refer to $\begin{bmatrix} m_1(\cdot) & m_2(\cdot) \end{bmatrix}'$). In general we have G-dimensional Y_{it} (and η_{it},ε_{it}) and K-dimensional X_{it}, where $G \leq K$. For notational simplicity the models in this chapter will have $G = K$. We restrict our focus on stationary models. That is to say that we assume $\{Y_{it}\}$ is a stationary Markov process.[3]

[2]This and other topics related to dynamic panel estimation (discrete choice models, etc.) are reserved for future research.

[3]This can be considered a restriction on the functions m but we leave the primitive assumptions assuring stationarity for future research.

2.1 A Two-Dimensional, Two Period Model

For illustrative purposes we first focus on a two equation model for determining outcomes in periods 0 and 1.

$$
\begin{aligned}
Y_{i1}^1 &= m_1(Y_{i1}^2, Y_{i0}^1, Y_{i0}^2, X_{i1}^1 + \varepsilon_{i1}^1), \qquad & \varepsilon_{i1}^1 &= h_1(\varepsilon_{i0}^1, \eta_{i1}^1) \\
Y_{i1}^2 &= m_2(Y_{i1}^1, Y_{i0}^1, Y_{i0}^2, X_{i1}^2 + \varepsilon_{i1}^2), \qquad & \varepsilon_{i1}^2 &= h_2(\varepsilon_{i0}^2, \eta_{i1}^2) \\
Y_{i0}^1 &= s_1(Y_{i0}^2, X_{i0}^1 + \varepsilon_{i0}^1) \\
Y_{i0}^2 &= s_2(Y_{i0}^1, X_{i0}^2 + \varepsilon_{i0}^2)
\end{aligned}
$$

where $X_{i0} \perp \varepsilon_{i0}$ and $X_{i1} \perp (\varepsilon_{i0}, \varepsilon_{i1})$.

These types of models do not have the triangular structure adopted by Imbens and Newey (2009) or Chesher (2003) nor do they conform to assumptions restricting the simultaneous equations models described by Matzkin (2008). To show identification of m and s, we adopt the intuition of Bhargava and Sargan (1983) and treat the entire system of equations as a larger set of simultaneous equations. As in the simultaneous equations models of Sargan (1961), Matzkin (2008), identification of the structural functions will depend on whether a rank condition holds, which depends on exclusion restrictions on the exogenous variables X. However, Sargan's results hold in the case of additively linear functions and Matzkin's are for cases without lagged dependent variables and autocorrelated unobservable variables. The solution can be described as a control function approach in which we make a conditional independence argument to solve the endogeneity of the lagged dependent variables.

There are a number of applications of these models. In the field of education policy, there is much research on the relationship between education expenditure and student achievement. Hanushek (1989) reviews the literature and states that there is little evidence of an impact of education expenditure on student achievement. One potential application for models of the type described in this paper are value-added models (VAM) of teacher quality. There are endogeneity problems: unobserved variables like student and teacher ability will be autocorrelated in addition to the simultaneous structure. Another application relates student achievement with teacher effort.[4] In the third chapter of this dissertation we discuss the application of this theory to estimating dynamic systems of supply and demand but restrict actual estimation to a single-equation model of demand for airline travel.

Abstracting from the issue of initial conditions, such models can take the form:

$$SA_t = m_1(SE_t, SA_{t-1}, SE_{t-1}, X_t^1 + \varepsilon_t^1) \qquad \varepsilon_t^1 = h_1(\varepsilon_{t-1}, \eta_t^1)$$

$$SE_t = m_2(SA_t, SE_{t-1}, SA_{t-1}, X_t^2 + \varepsilon_t^2) \qquad \varepsilon_t^2 = h_2(\varepsilon_{t-1}, \eta_t^2)$$

where SA is student achievement, SE is school expenditure (or teacher effort), and X_t are instrumental variables satisfying appropriate conditions.

[4] See Ballou and Podgursky (1995)

3 Identification

To show identification, we make the following assumptions:

Distributional Assumptions

We assume that

D1. η_{it} is independent of $(\varepsilon_{i,s-1})$, for $s \leq t$.

η_{it} are independently distributed over i.

X_{it} is independent of (ε_{is}) for $s \leq t$.

D2. The support of X_{it} is $\mathcal{R}^G$.

The support of $(\varepsilon_{i,t-1}, \eta_{it})$ is $\mathcal{R}^K \times \mathcal{R}^K$ (the densities are everywhere positive)

Densities of (X_{it}) are everywhere continuously differentiable.

Assumption D1 is on the relationship between the unobserved and observed variables. In each period, the innovation to the process for determining ε_{it}, η_t, is independent of prior and current regressors. A consequence of this assumption is that $Y_{i,t-1}$ and ε_{it} are independent conditional on $\varepsilon_{i,t-1}$. The rest of assumption D1 is a 'weak exogeneity' condition for X_{it} that allows for its use as an instrument.

Assumption D2 assures enough variation in X_{it} to cover the entire support of the unobservable variables. Also, the density of the unobservables is always positive; this ensures the monotonicity (and invertibility) of distribution functions. These support conditions can be loosened when considering identification of the model local to particular values.

Initial condition assumptions

I1. Exogenous initial conditions: In the initial period ($t = 0$),

$$Y_{i0}^1 = s_1(Y_{i0}^2, X_{i0}^1 + e_{i1}^1)$$

$$Y_{i1}^2 = s_2(Y_{i0}^1, X_{i0}^2 + e_{i1}^2)$$

Functional assumptions

F1. The vector valued function m is twice continuously differentiable and invertible in ε_{it}.

F2. For any (y_t, y_{t-1})

$$\left|\frac{\partial r(y_t, y_{t-1})}{\partial y_t}\right| > 0, \quad \text{where} \quad \frac{\partial r(y_t, y_{t-1})}{\partial y_t} \quad \text{is the Jacobian matrix for } r.$$

F3. X_{it}^g appears exactly in one of the G equations $m_1, \ldots, m_G$.

We use the invertibility of m to express the conditional density of $(Y_{i0}, \ldots, Y_{it})$ in terms of the density of $(\varepsilon_{it}, \varepsilon_{i,t-1})$, both conditional on $(X_{i0}, \ldots, X_{it})$. Assumption F2 is a sign normalization to ensure that densities have correct signs. Assumption F3 details exclusion restrictions on the vector of instruments X_{it}.

Theorem. *If the model is given by:*

$$Y_{it}^1 = m_1(Y_{it}^2, \ldots, Y_{it}^K, Y_{i,t-1}, X_{it}^1 + \varepsilon_{it}^1) \qquad \varepsilon_{it}^1 = h_1(\varepsilon_{i,t-1}^1, \eta_{it}^1)$$

$$\vdots$$

$$Y_{it}^K = m_K(Y_{it}^1, \ldots, Y_{it}^{K-1}, Y_{i,t-1}, X_{it}^K + \varepsilon_{it}^K) \qquad \varepsilon_{it}^K = h_K(\varepsilon_{i,t-1}^K, \eta_{it}^K)$$

$$\vdots$$

$$Y_{i0}^1 = s_1(Y_{i0}^2, \ldots, Y_{i0}^K, X_{i0}^1 + \varepsilon_{i0}^1)$$

$$\vdots$$

$$Y_{i0}^K = s_K(Y_{i0}^1, \ldots, Y_{i0}^{K-1}, X_{i0}^K + \varepsilon_{i0}^K)$$

Assumptions D, I, and F hold, and we observe the conditional density $f_{Y_t,Y_{t-1}|Y_{t-2},X_t,X_{t-1}}$*, then the derivatives of* m *are identified.*

For proving our theorem, we start by writing the density of $Y_i \equiv (Y_{iT}, \ldots, Y_{i0})$ conditional on $X_i \equiv (X_{iT}, \ldots, X_{i0})$ in terms of the density of $\varepsilon_i \equiv (\varepsilon_{iT}, \ldots, \varepsilon_{i0})$ conditional on X_i:

$$f_{Y_i}(y_i) = f_{\varepsilon_i|X_i}\left(r(y_{iT}, y_{i,T-1}) - x_{iT}, \ldots, r_0(y_{i0}) - x_{i0}\right)\left|\frac{\partial R(y_i)}{\partial y_i}\right|$$

Where R is the Jacobian of the vector-valued function made up of all the m functions. If $(r, f_{\varepsilon_i|x_i})$ and $(\tilde{r}, \tilde{f}_{\varepsilon_i|x_i})$ are observationally equivalent,

$$f_{\varepsilon_i|X_i}\left(r(y_{iT}, y_{i,T-1}) - x_{iT}, \ldots, r_0(y_{i0}) - x_{i0}\right)\left|\frac{\partial R(y_i)}{\partial y_i}\right|$$

$$=$$

$$f_{\tilde{\varepsilon}_i|X_i}\left(\tilde{r}(y_{iT}, y_{i,T-1}) - x_{iT}, \ldots, \tilde{r}_0(y_{i0}) - x_{i0}\right)\left|\frac{\partial \tilde{R}(y_i)}{\partial y_i}\right|$$

Following Matzkin (2008), we take logarithms of both sides and differentiate with respect to each y_{it}:

$$\frac{\partial \log f_{\varepsilon_i|x_i}}{\partial \varepsilon_{it}} \frac{\partial r(y_{it}, y_{i,t-1})}{\partial y_{it}} + \frac{\partial \log f_{\varepsilon_i|x_i}}{\partial \varepsilon_{i,t+1}} \frac{\partial r(y_{i,t+1}, y_{it})}{\partial y_{it}} + \frac{\partial \log\left(\left|\frac{\partial R}{\partial y_i}\right|\right)}{\partial y_{it}}$$

$$=$$

$$\frac{\partial \log f_{\tilde{\varepsilon}_i|x_i}}{\partial \tilde{\varepsilon}_{it}} \frac{\partial \tilde{r}(y_{it}, y_{i,t-1})}{\partial y_{it}} + \frac{\partial \log f_{\tilde{\varepsilon}_i|x_i}}{\partial \tilde{\varepsilon}_{i,t+1}} \frac{\partial \tilde{r}(y_{i,t+1}, y_{it})}{\partial y_{it}} + \frac{\partial \log\left(\left|\frac{\partial \tilde{R}}{\partial y_i}\right|\right)}{\partial y_{it}}$$

This generates T conditions that must hold for observational equivalence. We will show that the variation in X_i is sufficient to identify the derivatives of r and $\tilde{r}$.

Under additional assumptions we can also show that identification of the functions r imply the identification of the functions h, which govern the evolution of the unobservable variable ε_{it}. In our education applications, ε_{it} can carry the interpretation of quality (of students, teachers, institutions, etc.). [5] To identify the derivative of h with respect to $\varepsilon_{i,t-1}$, we make the following additional assumptions:

Additional Assumptions

[5] A function of this type, which described the development of cognitive and non-cognitive skill in students, is of primary interest for Cunha, Heckman and Schennach (2010).

U1. The function $r(y_{it}, y_{i,t-1}) = \gamma$ at a known set of values $(\bar{y}_{it}, \bar{y}_{i,t-1})$.

U2. The vector valued function h is twice continuously differentiable and invertible in η_{it}.

U3. For every $(\varepsilon_{it}, \varepsilon_{i,t-1})$

$$\left| \frac{\partial h^{-1}(\varepsilon_{i,t-1}, \varepsilon_{it})}{\partial \varepsilon_{it}} \right| > 0, \quad \text{where} \quad \frac{\partial h^{-1}(\varepsilon_{i,t-1}, \varepsilon_{it})}{\partial \varepsilon_{it}} \quad \text{is the Jacobian matrix for } h^{-1}.$$

U4. $f_{\eta_t}(n_t)$ is the same for all t.

Assumption U1 fixes the function r at a specific point in order to identify the function from the derivatives identified under assumptions D, I and F. Assumption U2 and U3 are the same as F2 and F3 and ensure that we can write the density of ε_{it} in terms of that of η_{it}.

Corollary.

If Assumption U holds in addition to the conditions in the above theorem, the derivative $\frac{\partial h(\varepsilon_{i,t-1},\eta_{it})}{\partial \varepsilon_{i,t-1}}$ *is identified.*

With the identification of the functions r, the values $e_t = r(y_t, y_{t-1}) - x_t$ are 'observable' for every t. Then the relationship

$$\varepsilon_t = h(\varepsilon_{t-1}, \eta_t)$$

is a set of nonparametric equations that can be estimated by a variety of methods, one of which is outlined by Matzkin (2003).[6]

[6]See appendix B for details.

Generalizing the model to multiple equations is also straightforward. The same intuition of Matzkin (2008) can be used - identification/non-identification will depend on whether a rank condition holds.

4 Conclusion

This chapter shows identification of a structural model with simultaneous equations where there is endogeneity due to correlation between lagged dependent variables and unobservable heterogeneity. Existing control function approaches would not be applicable because of the autocorrelation in the unobservable variables. We extend the results of Matzkin (2008) with a conditional independence argument in order to show identification of derivatives of the structural functions. This paper also makes a contribution in the estimation of nonparametric state dependence in the presence of serially correlated unobservable variables.

Future avenues for research include making closer ties with the dynamic panel literature; a more rigorous handling of initial conditions and accompanying treatment of permanent individual effects (fixed and random effects) would be helpful for making the model even more applicable. A number of the assumptions made in this chapter are somewhat strict - among them the exclusion restrictions on X_{it} and on $\varepsilon_{i,t-1}$ in the function $\varepsilon_{it} = h(\varepsilon_{i,t-1}, \eta_{it})$. A variety of generalizations of the function h (e.g., the h functions can have a simultaneous structure) are possible.

Appendix

If m, h and f_η are not restricted, m, h, and f_η are not separately identified, since changes in f_η can be 'undone' by corresponding changes in h and the same is true for m with respect to h. Therefore, we restrict m to be such that r is additively separable in X_{it}. This has the effect of normalizing the scale of ε_{it} to match that of X_{it}.[7]

$$\varepsilon_{it} = r(y_{it}, y_{i,t-1}, x_{it}) = r(y_{it}, y_{i,t-1}) - x_{it}$$

This form for r corresponds to these m functions:

$$Y^1_{it} = m_1(Y^2_{it}, Y^1_{i,t-1}, Y^2_{i,t-2}, X^1_{it} + \varepsilon^1_{it})$$

$$Y^2_{it} = m_2(Y^1_{it}, Y^1_{i,t-1}, Y^2_{i,t-2}, X^2_{it} + \varepsilon^2_{it})$$

4.1

To show identification of the derivatives of r (and thus, m, its inverse), we refer to the concept of 'observational equivalence.' We call functions (r, f_ε) and $(\tilde{r}, \tilde{f}_\varepsilon)$ "observationally equivalent" when either set could generate the observed density function $f_{Y_T,\ldots,Y_0|X_T,\ldots,X_0}$.

Under our assumptions we can write the density of $Y_i \equiv (Y_{iT}, \ldots, Y_{i0})$ conditional on

[7] Chesher (2003) makes a similar assumption by restricting (in our notation) $\frac{\partial m(Y_{i,t-1}, X_{it}, \varepsilon_{it})}{\partial \varepsilon_{it}} = 1$

$X_i \equiv (X_{iT}, \ldots, X_{i0})$ in terms of the density of $\varepsilon_i \equiv (\varepsilon_{iT}, \ldots, \varepsilon_{i0})$ conditional on X_i:

$$f_{Y_i}(y_i) = f_{\varepsilon_i|X_i}\left(r(y_{iT}, y_{i,T-1}) - x_{iT}, \ldots, r_0(y_{i0}) - x_{i0})\right)\left|\frac{\partial R(y_i)}{\partial y_i}\right|$$

Where R is the Jacobian of the vector-valued function made up of all the m functions. If $(r, f_{\varepsilon_i|x_i})$ and $(\tilde{r}, \tilde{f}_{\varepsilon_i|x_i})$ are observationally equivalent,

$$\begin{aligned} & f_{\varepsilon_i|X_i}\left(r(y_{iT}, y_{i,T-1}) - x_{iT}, \ldots, r_0(y_{i0}) - x_{i0}\right)\left|\frac{\partial R(y_i)}{\partial y_i}\right| \\ & \qquad = \\ & f_{\tilde{\varepsilon}_i|X_i}\left(\tilde{r}(y_{iT}, y_{i,T-1}) - x_{iT}, \ldots, \tilde{r}_0(y_{i0}) - x_{i0}\right)\left|\frac{\partial \tilde{R}(y_i)}{\partial y_i}\right| \end{aligned}$$

Following Matzkin (2008), we take logarithms of both sides and differentiate with respect to each y_{it}:

$$\begin{aligned} & \frac{\partial \log f_{\varepsilon_i|x_i}}{\partial \varepsilon_{it}}\frac{\partial r(y_{it}, y_{i,t-1})}{\partial y_{it}} + \frac{\partial \log f_{\varepsilon_i|x_i}}{\partial \varepsilon_{i,t+1}}\frac{\partial r(y_{i,t+1}, y_{it})}{\partial y_{it}} + \frac{\partial \log\left(\left|\frac{\partial R}{\partial y_i}\right|\right)}{\partial y_{it}} \\ & \qquad = \\ & \frac{\partial \log f_{\tilde{\varepsilon}_i|x_i}}{\partial \tilde{\varepsilon}_{it}}\frac{\partial \tilde{r}(y_{it}, y_{i,t-1})}{\partial y_{it}} + \frac{\partial \log f_{\tilde{\varepsilon}_i|x_i}}{\partial \tilde{\varepsilon}_{i,t+1}}\frac{\partial \tilde{r}(y_{i,t+1}, y_{it})}{\partial y_{it}} + \frac{\partial \log\left(\left|\frac{\partial \tilde{R}}{\partial y_i}\right|\right)}{\partial y_{it}} \end{aligned}$$

This condition is obviously different for periods T and 0. We also differentiate with respect to each x_{it}.

Adapting the notation in Matzkin (2008),

$$s_{\varepsilon_{it}} = \frac{\partial \log f_{\varepsilon_i|X_i}}{\partial \varepsilon_{it}},\ r_{y_{it}} = \frac{\partial r(y_{it}, y_{i,t-1})}{\partial y_{it}},\ r_{y_{it}^+} = \frac{\partial r(y_{i,t+1}, y_{it})}{\partial y_{it}},\ \Delta_{y_t} = \frac{\partial}{\partial y_{it}} \log \left| \frac{\partial r(y_{it}, y_{i,t-1})}{\partial y_{it}} \right|$$

the conditions for observational equivalence can be summarized:

$$\begin{aligned}
\tilde{s}_{\varepsilon_{iT}} \tilde{r}_{y_{iT}} + \tilde{\Delta}_{y_{iT}} &= s_{\varepsilon_{iT}} r_{y_{iT}} + \Delta_{y_{iT}} \\
\tilde{s}_{\varepsilon_{it}} \tilde{r}_{y_{it}} + \tilde{s}_{\varepsilon_{i,t+1}} \tilde{r}_{y_{it}^+} + \tilde{\Delta}_{y_{it}} &= s_{\varepsilon_{it}} r_{y_{it}} + s_{\varepsilon_{i,t+1}} r_{y_{it}^+} + \Delta_{y_{it}} \\
\tilde{s}_{\varepsilon_{i0}} \tilde{r}_{y_{i0}} &= s_{\varepsilon_{i0}} r_{y_{i0}} \\
\tilde{s}_{\varepsilon_{it}} &= s_{\varepsilon_{it}}
\end{aligned}$$

The condition with respect to y_{iT} (the first line) is:

$$\tilde{s}_{\varepsilon_{iT}} \tilde{r}_{y_{iT}} + \tilde{\Delta}_{y_{iT}} = s_{\varepsilon_{iT}} r_{y_{iT}} + \Delta_{y_{iT}}$$

or

$$\tilde{\Delta}_{y_{iT}} - \Delta_{y_{iT}} = [\tilde{s}_{\varepsilon_{iT}}] \tilde{r}_{y_{iT}} - [s_{\varepsilon_{iT}}] r_{y_{iT}}$$

Note that the last condition ensures that the terms in square brackets are equal. This means that

$$\tilde{\Delta}_{y_{iT}} - \Delta_{y_{iT}} = s_{\varepsilon_{iT}} \left(\tilde{r}_{y_{iT}} - r_{y_{iT}} \right)$$

The left side of this equation depends only on y_{iT} and $y_{i,T-1}$, as do the terms in the paranthe-ses). This means that if we fix y_{iT} and $y_{i,T-1}$ at particular values, as long as the middle term changes with $x_{it}, x_{i,t-1}$ or $y_{i,t-2}$, the only way for equality to hold is if both outside terms are zero. The middle term is $\frac{\partial \log f_{Y_i|X_i}}{\partial x_{it}}$. Thus, as long as the density of $(Y_{it}, Y_{i,t-1})$ changes with x_{it}, the derivative of r with respect to y_{it} is identified from the T period condition.[8]

The condition with respect to y_{it} (when t is not 0 or T) is:

$$[\tilde{s}_{\varepsilon_{it}}]\tilde{r}_{y_{it}} + [\tilde{s}_{\varepsilon_{i,t+1}}]\tilde{r}_{y_{it}^+} + \tilde{\Delta}_{y_{it}} = [s_{\varepsilon_{it}}]r_{y_{it}} + [s_{\varepsilon_{i,t+1}}]r_{y_{it}^+} + \Delta_{y_{it}}$$

Again, the conditions on x_{it} mean that the items in the square brackets are equal. Rearranging, we get that

$$\tilde{\Delta}_{y_{it}} - \Delta_{y_{it}} = s_{\varepsilon_{it}}\left[\tilde{r}_{y_{it}} - r_{y_{it}}\right] + s_{\varepsilon_{i,t+1}}[\tilde{r}_{y_{it}^+} - r_{y_{it}^+}]$$

From period T we determined that under our assumptions, $r_{y_{it}}$ is identified. The middle term is then zero and the same argument as above can be made for the remaining equation.

Matzkin (2008) shows that the above argument is equivalent to determining whether a rank condition on a particular matrix made up of the terms s_ε, r_y and Δ_y holds.

From the identification of derivatives of r, the derivatives of m (r's inverse) are also

[8]The argument for period 0 is almost exactly the same.

identified.

$$\frac{\partial m_1(y_t^2, y_{t-1}^1, y_{t-1}^2, x_t^1 + \varepsilon_t^1)}{\partial y_{t-1}^1} = \frac{\frac{\partial r(y_t, y_{t-1})}{\partial y_{t-1}^1}}{\frac{\partial r(y_t, y_{t-1})}{\partial y_t^1}}$$

4.2

With the derivatives of r and assumption A6, the entire function r is identified. This means that the realizations of ε_{it} are now 'observable' subject to the normalization made in assumption A6. We combine this 'observability' with the structure of the model to show identification of h, since

$$\begin{bmatrix} \varepsilon_{it}^1 \\ \vdots \\ \varepsilon_{it}^K \end{bmatrix} = \begin{bmatrix} h_1(\varepsilon_{i,t-1}^1, \eta_{it}^1) \\ \vdots \\ h_K(\varepsilon_{i,t-1}^K, \eta_{it}^K) \end{bmatrix}$$

Since r is identified, if we define $e_{it} = r(y_{it}, y_{i,t-1}) - x_{it}$ and $e_{i,t-1} = r(y_{i,t-1}, y_{i,t-2}) - x_{i,t-1}$,

$$\begin{bmatrix} e_{it}^1 \\ \vdots \\ e_{it}^K \end{bmatrix} = \begin{bmatrix} h_1(e_{i,t-1}^1, \eta_{it}^1) \\ \vdots \\ h_K(e_{i,t-1}^K, \eta_{it}^K) \end{bmatrix}$$

where e_{it} and $e_{i,t-1}$ are observable. For each $k \in 1, \ldots, K$, we have a nonparametric function of the form

$$e_{it}^k = h_k(e_{i,t-1}^k, \eta_{it}^k)$$

As part of A6 we also assume that for a particular value $\bar{\varepsilon}_{i,t-1}$, the function $h_k(\bar{\varepsilon}_{i,t-1}, \eta) = \eta$. With this assumption we can use the results of Matzkin (2003) to show that h_k is identified.

References

[1] Anderson, T.W. and C. Hsiao (1981). "Estimation of Dynamic Models with Error Components." *Journal of the American Statistical Association*, Vol. 76, No. 375.

[2] Ahn, S. and P. Schmidt (1991). "Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation." *Journal of Econometrics*, 76.

[3] Altonji, J. G., and R. L. Matzkin (2005). "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors." *Econometrica*, Vol. 73, No. 4.

[4] Arellano, M., and B. Honore (2001). "Panel Data Models: Some Recent Developments." *Handbook of Econometrics, Volume 5, J.J. Heckman and E. Leamer, eds.*.

[5] Ballou, D. and M. Podgursky (1995). "Education Policy and Teacher Effort." *Industrial Relations*, Vol. 34, No. 1.

[6] Baltagi, B. (1981). "Simultaneous Equations with Error Components." *Journal of Econometrics*, Vol. 17.

[7] Bester, C. A., and C. Hansen (2007). "Flexible Correlated Random Effects Estimation in Panel Models with Unobserved Heterogeneity." *University of Chicago Working Paper*.

[8] Bhargava, A., and J. D. Sargan (1983). "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods." *Econometrica*, Vol. 51, No. 6.

[9] Blundell R., and J. Powell (2003). "Endogeneity in Semiparametric and Nonparametric Regression Models." *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress*, Vol. 2.

[10] Chesher, A. (2003). "Identification in Nonseparable Models." *Econometrica*, Vol. 71, No. 5.

[11] Cornwell, C., Schmidt, P., and D. Wyhowski (1992). "Simultaneous equations and panel data." *Journal of Econometrics*, 51.

[12] Cunha, F., Heckman, J., and S. Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, Vol. 78, No. 3.

[13] Evdokimov, K. (2009). "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity." *Yale Working Paper.*

[14] Graham, B. and J. L. Powell (2010). "Identification and estimation of 'irregular' correlated random coefficient models." *Working Paper.*

[15] Haile, P., and S. Berry (2010). "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers." *Yale Working Paper.*

[16] Hannan, E. J. (1971). "The Identification Problem for Multiple Equation Systems with Moving Average Errors." *Econometrica*, Vol. 39, No. 5.

[17] Hanushek, E. A. (1989). "The Impact of Differential Expenditures on School Performance." *Educational Researcher*, Vol. 18, No. 4.

[18] Hatanaka, M. (1975). "On the Global Identification of the Dynamic Simultaneous Equations Model with Stationary Disturbances." *International Economic Review*, Vol. 16, No. 3.

[19] Hausman, J. (1983). "Specification and estimation of Simultaneous Equation Models." *Handbook of Econometrics, Vol. 1. Z. Griliches and M. D. Intrilligator, eds.*.

[20] Heckman, J. (1981). "Heterogeneity and State Dependence." *Studies in Labor Markets, Sherwin Rosen, editor.*

[21] Hoderlein, S., and H. White (2009). "Nonparametric identification in nonseparable panel data models with generalized fixed effects." *CEMMAP Working Paper*, CWP33/09.

[22] Hsiao, C. (1983). "Identification." *Handbook of Econometrics, Vol. 1. Z. Griliches and M. D. Intriligator, eds.*.

[23] Hsiao, C. (1986). "Analysis of Panel Data." *Econometric Society Monograph*, No. 11.

[24] Hu, Y., and M. Shum (2010). "Nonparametric Identification of Dynamic Models with Unobserved State Variables." *Johns Hopkins Working Paper.*

[25] Hu, Y., Shum, M., and W. Tan (2010). "A Simple Estimator for Dynamic Models with Serially Correlated Unobservables." *Johns Hopkins Working Paper*, 558.

[26] Hyslop, D. (1999). "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women." *Econometrica*, Vol. 67, No. 6.

[27] Imbens, G., and W. Newey (2009). "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity." *Econometrica*, Vol. 77, No. 5.

[28] Lee, Y. (2010). "Nonparametric Estimation of Dynamic Panel Models with Fixed Effects." *University of Michigan Working Paper.*

[29] Matzkin, R. L. (2003). "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica*, Vol. 71, No. 5.

[30] Matzkin, R. L. (2008). "Identification in Nonparametric Simultaneous Equations Models." *Econometrica*, Vol. 76, No. 5.

[31] Matzkin, R. L. (2010). "Estimation of Nonparametric Models with Simultaneity."

[32] Newey, W., Powell, J., and F. Vella (1999). "Nonparametric Estimation of Triangular Simultaneous Equations Models." *Econometrica*, Vol. 67, No. 3.

[33] Sargan, J. D. (1961). "The Maximum Likelihood Estimation of Econometric Relationships with Autoregressive Residuals." *Econometrica*, Vol. 29, No. 3.

[34] Sargan, J. D. "The Identification and Estimation of Sets of Simultaneous Stochastic Equations." *Contributions to Econometrics: John Denis Sargan*, Vol. 1.

[35] Schennach, S., White, H., and K. Chalak (2009). "Local Indirect Least Squares and Average Marginal Effects in Nonseparable Structural Systems." *UCSD Working Paper*.

[36] Shiu, J.L., and Y. Hu (2010). "Identification and Estimation of Nonlinear Dynamic Panel Data Models with Unobserved Covariates." *Johns Hopkins Working Paper*, 557.

[37] Su, L., and A. Ullah (2010). "Nonparametric and Semiparametric Panel Econometric Models: Estimation and Testing." *Singapore Management University Working Paper*.

Nonparametric Estimation of Dynamic Simultaneous Equations Models

1 Introduction

In this chapter we build on the discussion of identification of nonparametric dynamic simultaneous equations models from the previous chapter and describe methods for estimation. The estimators are constructed to correct for two sources of endogeneity. The first is due to the simultaneous equations structure and the second is due to the inclusion of lagged dependent variables as regressors. These estimators are shown to be consistent and asymptotically normal. In order to alleviate the 'curse of dimensionality' that is endemic to nonparametric estimation, we use the techniques shown by Matzkin (2010) to 'average over' each of the exogenous variables. Even with this reduction in dimension the conclusion is that for most applications, semiparametric estimation would be the preferred choice. We present a limited view of some small-sample properties of these estimators using simulations. The simulations highlight some of the practical drawbacks to nonparametric estimation.

In the previous chapter we discussed how dynamic nonparametric simultaneous equations models with G equations in T time periods can be characterized as a set of $G \times T$ simultaneous equations as described by Bhargava and Sargan (1983) for the linear case. This allows us to apply estimation techniques already developed for 'standard' simultaneous equations models. The drawback to this is the rapid increase in the size of the model for moderate increases in the number of time periods and/or number of equations. Not only is there a curse of dimensionality in the convergence rate of our estimators, in terms of the practical implementation of the estimator, things like cross-validation and even the selection of points over which to estimate the model grow very quickly. Despite these caveats, since nonparametric dynamic models with endogenous regressors are increasingly economically important, it is useful to understand how to conduct estimation.

As stated in the previous chapter, simultaneous equations models with dynamic elements have been of interest for decades, but exclusively in linear form. Sargan (1961) focused on identification issues from such models, and a number of subsequent papers were written by Fair (1972), Hatanaka (1974) and Fomby (1984). These papers discuss efficient estimation of these models using panel data.

The purpose of the research in this dissertation is to take these well-studied models and relax the assumption of linearity by using recent advances in the estimation of nonparametric simultaneous equations models. This chapter is closely related to that of Matzkin (2010), which describes an estimation method for a system of simultaneous equations in one period. Similar to Matzkin's earlier papers on single-equation nonparametric estimation, the estimation is based on the identification of derivatives of the structural functions. In Matzkin (2010) the structure of the equations and their derivatives lends itself to a least-squares type estimator. The work in this dissertation expands upon the results from Matzkin (2008) to dynamic-panel type models; the reason to characterize the model as it is in the previous chapter is so that, for estimation, methods from the well-studied 'static' simultaneous equations model are applicable.

Naturally, the techniques discussed in this chapter relate to the estimation of nonparametric/nonseparable models generally. In particular, the estimation of dynamic models is closely related to the estimation of triangular simultaneous models, since identification and estimation is achieved through control variable techniques - the lagged dependent variables can be conditional independent given certain assumptions about initial conditions. Chesher (2003) and Imbens and Newey (2009) do not explicitly mention dynamic models, but careful application of their estimation techniques can be done for nonseparable models with dynamics.

One benefit of using the Matkzin (2010) technique is that other methods require inverting moments. When done with nonlinear models, this leads to the "ill-posed inverse" problem. This requires the imposition of completeness conditions on the distributions of unobservable variables. Imbens and Newey (2009) is one well-known example of an estimation method that requires trimming or resorting to other methods to deal with the ill-posed inverse problem.

We first briefly discuss the issues of identification in dynamic nonparametric models with simultaneous equations (the full treatment is in the previous chapter). We follow with a section describing the form of the proposed estimators. Section 3 discusses the asymptotic distribution and the results of the simulations of the model. Section 4 concludes.

2 Identification

A typical supply and demand model:

$$Q_{i1} = \gamma_1 P_{i1} + \rho_1 Q_{i0} + \varepsilon_{i1}^1$$

$$P_{i1} = \gamma_2 Q_{i1} + \rho_2 Q_{i0} + \varepsilon_{i1}^2$$

The identification argument in the first chapter establishes the link between nonparametric dynamic simultaneous equation models with previous work on linear dynamic panel models. In both, there are systems of equations in each period, and the key insight, as found in a paper by Bhargava and Sargan (1987) is 'expanding' the linear dynamic panel (in their case with only one equation per period) to become a ($T \times 1$) set of simultaneous equations. Extending this logic, we can say that a dynamic panel model with G simultaneous equations in each period could be thought of as a $T \times G$ set of simultaneous equations. The identification then uses existing results on nonparametric systems of simultaneous equations applied to the larger set of estimating equations.

Our identification strategy uses ratios of derivatives of conditional probability density functions of observable variables. Even with structural functions that are nonseparable between variables, we show that particular items of interest, like marginal effects (derivatives of the structural function) are identified from linear combinations of these derivatives. The constructive nature of the identification arguments means that we can easily replace these functions of conditional density functions with their sample analogs from data. The procedure is straightforward and is a direct application of existing estimators for nonparametric simultaneous equations models.

For practitioners, the message of this chapter will most likely be that semiparametric estimation (instead of 'fully' nonparametric estimation) is prudent for a number of theoretical and practical reasons detailed below. However, identification and the principles of estimating the nonparametric model are still valuable. In general, the nonparametric model almost always correctly specifies the true model (i.e., there are no structural functions relating the specified variables that are not special cases of the nonparametric model) while semiparametric (or linear) versions of the model could potentially suffer from model misspecification. It is our opinion, therefore, that attention to the nonparametric identification of estimation models is important even when there is no intention to use nonparametric estimation methods.[9]

Our assumptions preclude the inclusion of 'fixed effects' models as special cases. This is a potential source of bias, since the true model may have fixed effects; that is, permanent unobservable characteristics or conditions that are related to variables in each period. Innovations in econometric theory or imposition of more structure could lead to identification and estimation of nonparametric/nonseparable dynamic panel fixed effects models. With such advancements, the potential for Hausman-type tests for the exogeneity of covariates also exists. We leave these topics for future research.

[9]Future research on developing nonparametric specification and omitted variable tests (in the style of Fan and Li (1996)) specifically for use with dynamic nonparametric simultaneous equations models might be of particular interest for applied researchers.

3 Estimation

As discussed in the introduction to this chapter, one significant benefit to casting the dynamic problem as a set of nonparametric simultaneous equations that estimation is a straightforward application of Matzkin (2010). The method produces an $(X'X)^{-1}(X'Y)$-like estimator. As in the discussion of identification, the difference between the models of Matzkin (2010) are the state dependence and autocorrelated errors, which means that Y_0 will have two effects on Y_1 - one direct effect and one through the autocorrelation of ε.

Again, the model is:

$$
\begin{aligned}
&Y^1_{iT} = m_1(Y^2_{iT}, Y^1_{i,T-1}, Y^2_{i,T-1}, X^1_{iT} + \varepsilon^1_{iT}) \qquad \varepsilon^1_{iT} = h_1(\varepsilon^1_{i,T-1}, \eta^1_{iT})\\
&Y^2_{iT} = m_2(Y^1_{iT}, Y^1_{i,T-1}, Y^2_{i,T-1}, X^2_{iT} + \varepsilon^2_{iT}) \qquad \varepsilon^2_{iT} = h_2(\varepsilon^2_{i,T-1}, \eta^2_{iT})\\
&\quad\vdots\\
&Y^1_{it} = m_1(Y^2_{it}, Y^1_{i,t-1}, Y^2_{i,t-1}, X^1_{it} + \varepsilon^1_{it}) \qquad \varepsilon^1_{it} = h_1(\varepsilon^1_{i,t-1}, \eta^1_{it})\\
&Y^2_{it} = m_2(Y^1_{it}, Y^1_{i,t-1}, Y^2_{i,t-1}, X^2_{it} + \varepsilon^2_{it}) \qquad \varepsilon^2_{it} = h_2(\varepsilon^2_{i,t-1}, \eta^2_{it})\\
&\quad\vdots\\
&Y^1_{i0} = s_1(Y^2_{i0}, X^1_{i0} + \varepsilon^1_{i0})\\
&Y^2_{i0} = s_2(Y^1_{i0}, X^2_{i0} + \varepsilon^2_{i0})
\end{aligned}
$$

Since the paper by Matkzin (2010) focuses on the estimation of the simultaneous equations, we focus specifically on the estimation of the lagged dependent variable's effect on current variables.

The invertibility conditions in assumption F allow us to write the density function of (Y_{i1}, Y_{i0}) conditional on (X_{i1}, X_{i0}) as a function of $(\varepsilon_{i1}, \varepsilon_{i0})$:

$$f_{Y_T,\ldots,Y_0|X_T,\ldots,X_0}(y_T,\ldots,y_0;x_T,\ldots,x_0) = f_{\varepsilon_T,\ldots,\varepsilon_0|X_T,\ldots,X_0}(r,d)\left|\frac{\partial R}{\partial(y_T,\ldots,y_0)}\right|$$

where r and d are the inverse functions of m and s. Taking the logarithm of both sides and taking derivatives (thanks to the smoothness conditions on the structural and density functions), we get the following conditions:

$$\begin{aligned}
q_{y_{iT}} &= s_{\varepsilon_{iT}}\frac{\partial r(y_{iT}, y_{i,T-1})}{\partial y_{iT}} + \Delta_{y_{iT}} \\
q_{y_{it}} &= s_{\varepsilon_{it}}\frac{\partial r(y_{it}, y_{i,t-1})}{\partial y_{it}} + s_{\varepsilon_{i,t+1}}\frac{\partial r(y_{i,t+1}, y_{it})}{\partial y_{it}} + \Delta_{y_{it}} \\
q_{x_{it}} &= s_{\varepsilon_{it}}
\end{aligned}$$

Where

$$q_{y_t} = \frac{\partial \log f_{Y_i|X_i}}{\partial y_{it}},\ s_{\varepsilon_{it}} = \frac{\partial \log f_{\varepsilon_i|X_i}}{\partial \varepsilon_{it}},\ \Delta_{y_{it}} = \frac{\partial}{\partial y_{it}}\log\left|\frac{\partial r(y_{it}, y_{i,t-1})}{\partial y_{it}}\right|$$

For notational simplicity, if we had a two period model ($t = 0, 1$), the conditions would be:

$$
\begin{aligned}
q_{y_1} &= q_{\varepsilon_1}\frac{\partial r(y_1,y_0)}{\partial y_1} + \Delta_{y_1} \\
q_{y_0} &= q_{\varepsilon_1}\frac{\partial r(y_1,y_0)}{\partial y_0} + q_{\varepsilon_0}\frac{\partial r_0(y_0)}{\partial y_0} + \Delta_{y_0} \\
q_{x_1} &= q_{\varepsilon_1}, \qquad q_{x_0} = q_{\varepsilon_0}
\end{aligned}
$$

The estimator for $\frac{\partial r(y_1,y_0)}{\partial y_1}$ differs from that in Matzkin (2010) because it explicitly depends on y_0, but the form of the estimator is the same.

To estimate $\frac{\partial m(y_1,y_0,x_1+e_1)}{\partial y_0}$, take the derivatives with respect to Y_{i0}:

$$
q_{y_0} = q_{\varepsilon_1}\frac{\partial r(y_1,y_0)}{\partial y_0} + q_{\varepsilon_0}\frac{\partial d_0(y_0)}{\partial y_0} + \Delta_{y_0} \tag{1}
$$

Integrating with respect to $x = (x_0, x_1)$ and weighted by a function $\mu(x)$,

$$
\int q_{y_0}\mu(x)dx = \left[\int q_{x_1}\mu(x)dx\right]\frac{\partial r(y_1,y_0)}{\partial y_0} + \left[\int q_{x_0}\mu(x)dx\right]\frac{\partial d_0(y_0)}{\partial y_0} + \Delta_{y_0}
$$

where $\mu(x)$ is a bounded and continuous function that takes a value zero outside a compact set.

Subtracting the integrated condition from the original condition (equation (1)),

$$
q_{y_0} - \int q_{y_0} = \left[q_{x_1} - \int q_{x_1}\right]\frac{\partial r(y_1,y_0)}{\partial y_0} + \left[q_{x_0} - \int q_{x_0}\right]\frac{\partial d_0(y_0)}{\partial y_0}
$$

where $\int q_{y_0} \equiv \int q_{y_0}\mu(x)dx$.

Defining $Q_{y_0^1} \equiv q_{y_0^1} - \int q_{y_0^1}$, we can rewrite the conditions in matrix form:

$$\begin{aligned}
\begin{bmatrix} Q_{y_0^1} & Q_{y_0^2} \end{bmatrix} &= \begin{bmatrix} Q_{x_1^1} & Q_{x_1^2} \end{bmatrix} \frac{\partial r(y_0, y_1)}{\partial y_0} + \begin{bmatrix} Q_{x_0^1} & Q_{x_0^2} \end{bmatrix} \frac{\partial d(y_0)}{\partial y_0} \\
&= \begin{bmatrix} Q_{x_1^1} & Q_{x_1^2} & Q_{x_0^1} & Q_{x_0^2} \end{bmatrix} \begin{bmatrix} \frac{\partial r(y_0,y_1)}{\partial y_0} \\ \frac{\partial d(y_0)}{\partial y_0} \end{bmatrix}
\end{aligned}$$

Now, (left) multiplying both sides of the condition by $\begin{bmatrix} Q_{x_1^1} & Q_{x_1^2} & Q_{x_0^1} & Q_{x_0^2} \end{bmatrix}'$, we can get the following system of equations:

$$\underbrace{\begin{bmatrix} Q_{x_1^1}Q_{y_0^1} & Q_{x_1^1}Q_{y_0^2} \\ Q_{x_1^2}Q_{y_0^1} & Q_{x_1^2}Q_{y_0^2} \\ Q_{x_0^1}Q_{y_0^1} & Q_{x_0^1}Q_{y_0^2} \\ Q_{x_0^2}Q_{y_0^1} & Q_{x_0^2}Q_{y_0^2} \end{bmatrix}}_{\text{``}\Gamma_{y_1,y_0}(y_1,y_0)\text{''}} = \underbrace{\begin{bmatrix} Q_{x_1^1}Q_{x_1^1} & Q_{x_1^1}Q_{x_1^2} & Q_{x_1^1}Q_{x_0^1} & Q_{x_1^1}Q_{x_0^2} \\ Q_{x_1^2}Q_{x_1^1} & Q_{x_1^2}Q_{x_1^2} & Q_{x_1^2}Q_{x_0^1} & Q_{x_1^2}Q_{x_0^2} \\ Q_{x_0^1}Q_{x_1^1} & Q_{x_0^1}Q_{x_1^2} & Q_{x_0^1}Q_{x_0^1} & Q_{x_0^1}Q_{x_0^2} \\ Q_{x_0^2}Q_{x_1^1} & Q_{x_0^2}Q_{x_1^2} & Q_{x_0^2}Q_{x_0^1} & Q_{x_0^2}Q_{x_0^2} \end{bmatrix}}_{\text{``}\Pi_{x_1,x_0}(y_1,y_0)\text{''}} \begin{bmatrix} \frac{\partial r(y_0,y_1)}{\partial y_0} \\ \frac{\partial d(y_0)}{\partial y_0} \end{bmatrix}$$

We integrate both sides over the function $\mu(x)$ and obtain our estimator by replacing the matrices Π and Γ with their sample analogs:

$$\begin{bmatrix} \frac{\partial r(y_1,y_0)}{\partial y_0} \\ \frac{\partial d(y_0)}{\partial y_0} \end{bmatrix} = [\Pi_{x_1,x_0}(y_1, y_0)]^{-1} [\Gamma_{y_1,y_0}(y_1, y_0)]$$

Again following the notation of Matzkin (2010), we analyze the asymptotic distribution of $\frac{\partial \hat{r}(y_1,y_0)}{\partial y_0}$ by first defining $rr(y_1, y_0) \equiv vec\left(\begin{bmatrix} \frac{\partial r(y_1,y_0)}{\partial y_0} \\ \frac{\partial d(y_0)}{\partial y_0} \end{bmatrix}\right)$, $TT_{xx}(y_1, y_0) \equiv I_G \otimes \Pi_{x_1,x_0}$

Again, a big advantage to characterizing the model as a larger set of simultaneous equations is the existence of previous work to derive the asymptotic distribution. In order to prove consistency and asymptotic normality of our estimators for $\frac{\partial r(y_{it},y_{i,t-1})}{\partial y_{it}}$ and $\frac{\partial r(y_{it},y_{i,t-1})}{\partial y_{i,t-1}}$, we make the following assumptions (adapted from work by Newey (1994) on partial means estimators and Matzkin (2010)).

3.0.1 Assumption E

E1. $K(u)$ is a second order kernel. K is zero outside of a compact set, integrates to 1 and is differentiable of order $\Delta \geq 2$. These Δ derivatives are Lipschitz continous.

E2. f_{Y_i,X_i} is at least four times continuously differentiable

E3. $\mu(x)$ is bounded, continuous almost everywhere and zero outside a compact set τ.

E4. The bandwidth σ is such that the following hold:

1. $\sigma \to 0$
2. $N\sigma^{TG+2} \to \infty$:
3. $\sqrt{N}\sigma^{GT/2+1+s} \to 0$:
4. $\frac{N\sigma^{2TG+2}}{\ln N} \to \infty$:
5. $(\sqrt{N}\sigma^{TG/2+1})(\sqrt{\ln N/N\sigma^{2TG+2}} + \sigma^2)^2 \to 0$:

Under assumption ‘E’ the limit distribution of $\hat{rr}(y_1, y_0)$ has the same form as that of the static simultaneous equations estimator:

$$\sqrt{N\sigma}(\hat{rr}(y_1, y_0) - rr(y_1, y_0)) \to \left(0, TT_{xx}(y_1, y_0)^{-1}V(y_1, y_0)TT_{xx}(y_1, y_0)^{-1}\right)$$

$$\text{where} \qquad V(y_1, y_0) = \int \mu(x) \left\{ \int \tilde{K}(\tilde{y}, x) \tilde{K}(\tilde{y}, x)' d\tilde{y} d\tilde{x} \right\} \mu(x)' f_{Y,X} dx$$

The derivation of the asymptotic variance is due to the 'partial mean' interpretation of the estimator $\frac{\partial \hat{r}(y_1,y_0)}{\partial y_0}$. Note that the rate of convergence is related only to the number of endogenous variables.

To isolate $\frac{\partial r(y_1,y_0)}{\partial y_0}$,

$$rr^0(y_1, y_0) \equiv vec\left(\frac{\partial r(y_1, y_0)}{\partial y_0}\right) = I_2 \otimes \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} [rr(y_1, y_0)]$$

As seen in the convergence rate of the estimators, the multi-period model suffers doubly from the 'curse of dimensionality' as the time dimension increases, since the dimension scales with $T \times G$ and not just G. For this reason the practical choice will often have to be semiparametric estimation.

4 Simulations

In order to assess the small sample properties of our estimation methods, we conducted a limited Monte Carlo study using data simulated from a nonseparable model. The data are simulated loosely based on a model of utility maximization where N individuals choose, over two periods, how much of two goods (y^1 and y^2) to consume. Just as in the previous section, we focuse on the effect of the Y_0s on the Y_1s.

In the initial period, (period 0), individuals maximize:

$$U(y_0^1, y_0^2) = (y_0^1)^\alpha (y_0^2)^\beta - \varepsilon_0^1 y_0^1 - \varepsilon_0^2 y_0^2$$

$$s.t. \quad p_0^1 y_0^1 + p_0^2 y_0^2 = I_0$$

In the second period, given the choices of period 0, individuals maximize:

$$U(y_0^1, y_0^2, y_1^1, y_1^2) = \left[(y_0^1)^{\alpha_1}(y_1^1)^{\alpha_2}\right]^\alpha \left[(y_0^2)^{\alpha_3}(y_1^2)^{\alpha_4}\right]^\beta - \varepsilon_1^1 y_1^1 - \varepsilon_1^1 y_1^1$$

$$s.t. \quad p_1^1 y_1^1 + p_1^2 y_1^2 = I_1$$

Since the first order conditions determining the solution of this model are:

$$\varepsilon_1^1 = \left[(y_0^1)^{\alpha 1 \alpha}(y_0^2)^{\alpha_3 \beta}\right] \alpha_2 \alpha y_1^{\alpha_2 \alpha - 1} y_1^{\alpha_4 \beta} - p_1^1$$

$$\varepsilon_1^2 = \left[(y_0^1)^{\alpha 1 \alpha}(y_0^2)^{\alpha_3 \beta}\right] \alpha_4 \beta y_1^{\alpha_2 \alpha} y_1^{\alpha_4 \beta - 1} - p_1^2$$

using the notation of this paper,

$$r_1(y_0, y_1) = \left[(y_0^1)^{\alpha 1 \alpha}(y_0^2)^{\alpha_3 \beta}\right] \alpha_2 \alpha y_1^{\alpha_2 \alpha - 1} y_1^{\alpha_4 \beta}$$

$$r_2(y_0, y_1) = \left[(y_0^1)^{\alpha 1 \alpha}(y_0^2)^{\alpha_3 \beta}\right] \alpha_4 \beta y_1^{\alpha_2 \alpha} y_1^{\alpha_4 \beta - 1} - p_1^2$$

We simulate 100 samples of $N = 100$, 1000 and 5000 to assess the (and in doing so, illustrate some small-sample properties of estimating the entire system).[10]

[10] The code to perform these simulations, which includes the parameter values used, is available at

We also ran estimation on 100 samples of data generated through a linear simultaneous equation model.

4.1 Simulation results

When Truth is Nonseparable:

		$N = 500$		$N = 1000$		$N = 2500$	
	(True Value)	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\frac{\partial r_1(y_0,y_1)}{\partial y_0^1}$	-20.6913	16.2198	24.6918	11.174	23.4973	-2.3123	12.6987
$\frac{\partial r_1(y_0,y_1)}{\partial y_0^2}$	-20.6913	19.4723	26.0850	14.8608	24.0841	3.8346	12.6414
$\frac{\partial r_2(y_0,y_1)}{\partial y_0^1}$	-20.6913	20.7138	30.0548	14.2784	25.6900	5.3452	14.6323
$\frac{\partial r_2(y_0,y_1)}{\partial y_0^2}$	-20.6913	16.9598	27.1661	11.3998	24.3153	-1.1912	15.2334

When True Model is Linear:

		$N = 500$		$N = 1000$		$N = 2500$	
	(True Value)	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\frac{\partial r_1(y_0,y_1)}{\partial y_0^1}$	1	-0.9393	3.1977	-0.6800	2.2011	-1.2125	2.8403
$\frac{\partial r_1(y_0,y_1)}{\partial y_0^2}$	0.8	-0.7760	4.4775	-0.5152	4.1716	-0.8022	3.7665
$\frac{\partial r_2(y_0,y_1)}{\partial y_0^1}$	0.3	-0.4263	3.1941	-1.0221	4.6274	-0.2178	3.2038
$\frac{\partial r_2(y_0,y_1)}{\partial y_0^2}$	1.2	-0.8249	2.7245	-0.9957	3.3496	-0.4299	2.6011

The simulations results show that when the true model is linear, the nonparametric

http://dmkang.bol.ucla.edu/research.html estimator is biased toward zero and very loosely estimated. However, when the true model is nonlinear (and nonseparable), the (slow) convergence to the true values is apparent. While linear models are precisely estimated, they would be very wrong and the results would have no structural interpretation.

4.2 Discussion

The results of the simulation underscore the advice for practitioners that 'fully' nonparametric estimation of these models requires a great amount of data and, as a result, of computing power as well. For instance, for feasibility reasons the derivatives of the (inverse) structural function were only calculated at 3 points per dimension of Y (total of $3^4 = 81$ points) and 5 points per dimension of X (625 points). In practice the estimator is likely better behaved when calculated at more points.

The warnings on computing power only continue due to the fact that our bandwidth selection for these simulations are likely also sub-optimal. Bandwidths were chosen according to the minimization of the cross-validation criterion for conditional density estimation (according to Fan and Yim (2004)). However, cross-validation was not performed for each simulation - instead we crudely 'estimated' the constant term for a rule-of-thumb based bandwidth that remained the same over each set of simulations.

These warnings underscore the difficulty of applying nonparametric estimation in general, but particularly for these models since the dimension of the model is large to begin with and increases quickly. Even in our illustration model with two periods and two equations, the theoretical convergence rate is on the order of $N^{-1/6}$ and the practical estimation is fraught with steps that can take a large amount of time.

With all that said, the research undertaken thus far has revealed several avenues by which the reliability of estimators can be increased from a practical standpoint. For instance, there are a number of data-driven ways to choose gridpoints for estimation, and there are also ways we can introduce cross-validation type intuition to select not only bandwidths but dimensions and spacing for gridpoints. These considerations, driven by necessity, might lead to advancements in the way nonparametric estimation is conducted for a wide variety of models. Of course, advancements in computing power and the ever-growing availability of data (particularly panel data) may eventually render these practical concerns obsolete.

5 Conclusion

In this chapter we presented an estimator for a system of dynamic nonparametric simultaneous equations. We related the identification from the first chapter of this dissertation to the construction of an estimator based on existing nonparametric estimators of simultaneous equations models, particularly the class of estimators proposed by Matzkin (2010). These estimators are consistent and asymptotically normal. However, one conclusion of this chapter is that the curse of dimensionality dramatically reduces the usability of these estimators unless large amounts of data and computing resources are available. The value in nonparametric estimation, of course, is the generality of the specification - though dimensionality makes estimation computationally burdensome, there is potentially (severe) bias from misspecification of the model.

Further research into issues of estimation of dynamic nonparametric simultaneous equations models will be focused on ways, as discussed in the immediately previous section, to improve its implementation. A more comprehensive study that incorporates optimal bandwidth/kernel selection is also of keen interest. More generally we would like to construct a guide for other researchers to formally test their parametric restrictions and to give them options as to how to conduct their estimation in the most robust yet efficient (using multiple meanings of the word) manner possible.

References

[1] Bhargava, A., and J. D. Sargan (1983). "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods." *Econometrica*, Vol. 51, No. 6.

[2] Chesher, A. (2003). "Identification in Nonseparable Models." *Econometrica*, Vol. 71, No. 5.

[3] Fair, R. (1972). "Efficient Estimation of Simultaneous Equations with Auto-regressive Errors by Instrumental Variables." *The Review of Economics and Statistics.*

[4] Fan, J. and T. H. Yim (2004). "A crossvalidation method for estimating conditional densities." *Biometrika*, Vol. 91, No. 4.

[5] Fomby, T. (1984). "Small sample efficiency gains from a first observation correction for Hatanaka's estimator of the lagged dependent variable-serial correlation regression model." *Federal Reserve Bank of Dallas Working Paper.*

[6] Hatanaka, M. (1974). "An Efficient Two-Step Estimator for the Dynamic Adjustment Model with Autoregressive Error." *Journal of Econometrics*, Vol. 2.

[7] Hatanaka, M. (1975). "On the Global Identification of the Dynamic Simultaneous Equations Model with Stationary Disturbances." *International Economic Review*, Vol. 16, No. 3.

[8] Imbens, G., and W. Newey (2009). "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity." *Econometrica*, Vol. 77, No. 5.

[9] Lee, Y. (2010). "Nonparametric Estimation of Dynamic Panel Models with Fixed Effects." *University of Michigan Working Paper.*

[10] Matzkin, R. L. (2003). "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica*, Vol. 71, No. 5.

[11] Matzkin, R. L. (2008). "Identification in Nonparametric Simultaneous Equations Models." *Econometrica*, Vol. 76, No. 5.

[12] Matzkin, R. L. (2010). "Estimation of Nonparametric Models with Simultaneity." *UCLA Working Paper*.

[13] Newey, W. (1994). "Kernel estimation of Partial means and a General Variance Estimator." *Econometric Theory,* Vol. 10.

[14] Sargan, J. D. (1961). "The Maximum Likelihood Estimation of Econometric Relationships with Autoregressive Residuals." *Econometrica*, Vol. 29, No. 3.

Nonparametric Estimation of Dynamic Demand Functions for Airline Travel

1 Introduction

In this chapter we estimate dynamic demand for airline travel using methods from nonparametric econometrics. The purpose is to relax assumptions on the demand function - in particular, the linear, additively separable relationship between price and quantity. Linearity means that the elasticity of demand is independent of the level of price. As discussed in the previous two chapters, the assumption of linearity brings many advantages in estimation but also brings the risk of misspecification. In this application, there is evidence that the price elasticity of airline travel changes with price. As a result, estimating price elasticities using regression models that are linear (or log-linear) would be incorrect. Policymakers and those responsible for pricing strategy could be basing their decisions on estimates with no structural interpretation. Relaxing parametric assumptions in these models gives us estimates of elasticities with bigger variance but with robustness with respect to incorrectly specifying the model.

There are a number of reasons why we expect the price elasticity of demand not to be constant over the range of prices for airline tickets. Anecdotal evidence suggests that, when shopping for airline tickets, people are more or less price-sensitive dependending on the type of travel. Indeed, in many of the studies cited later in this chapter, there is a distinction made between short-haul and long-haul travel and between business and leisure travel; when samples are separated by type of travel, estimated elasticities be very different. This evidence suggests that elasticities are price dependent, or at least not constant over the entire support of prices.

There are also theoretical reasons why we would expect demand to be price-dependent. In a recent paper, Jaffe and Weyl (2010) show that if individuals are basing their decisions according to typical discrete choice models, the aggregated market demand function will not be additively separable between prices. As they describe it, linear demand requires the (percentage) change in demand to be independent of the level of price. However, cheaper airlines have more consumers to begin with. If all prices increase, these cheaper airlines will have more consumers who are eligible to switch away and the elasticity of demand will be price-dependent. Our estimation results do - changing prices changes demand differently depending on the range of prices that an airline offers and is also not additively separable in competitors' prices.

Ironically, after spending most of this dissertation discussing dynamic simultaneous equations models, we will, in this chapter, not estimate the entire system of equations and instead use proxies for endogenous variables. We describe elasticities in a general sense, because changes in these variables will not correspond one-to-one with changes in price. This is one advantage of linear models - instrumental variable and control function methods are much more easily applicable.

The data for this chapter are collected from the DB1B database collected by the US Department of Transportation. The database is a 10% sample of all tickets sold for commercial airline travel within the United States. Demand is measured in 'revenue-passenger-miles,' which is the product of miles traveled by each passenger and the number of passengers.[11] We also have data on price paid for each itinerary in the sample. We construct a panel dataset to nonparametrically estimate dynamic demand equations.

We first highlight the economic literature on the demand for air travel and nonparametric estimation in Section 2. Section 3 discusses our various estimation models and the theoretical basis for estimating demand functions with non-constant elasticities. We then describe our data and instruments for price in Section 4. The results of our estimation are found in Section 5. We conclude this chapter and dissertation in Section 6. A brief appendix describing more details on our data follows.

2 Literature

The relevant literature covers the estimation of dynamic demand functions and nonparametric estimation. Naturally, the first estimation techniques developed for estimating demand equations were linear. More recent work focuses on the estimation of nonlinear and nonparametric demand functions, for instance papers by Kelejian (1971) and Roehrig (1988). Later papers expanded estimation to include nonlinear specifications of demand. Nonlinear demand functions are now used in a variety of applications. Two examples are Schmalansee

[11] 100 passengers traveling 100 miles would be traveling 10,000 revenue-passenger-miles

and Stoker (1999) and Blundell, Horowitz and Parey (2011); both papers are about estimating the demand for gasoline.

The airline industry, by itself, has had particular focus from economists due to the large amounts of publicly available data and the relative economic importance of air travel. The demand for airline travel has been studied by, among many others, Choi (1991), McGuire and Staelin (1983), and Oum, Waters and Yong (1990). Oum, et al's (1990) paper is relevant to this chapter because it summarizes the early research done on price elasticities in transportation. The more recent research on the economics of the airline industry has been focused on structural models of discrete choice, dynamic games between airlines, and other competitive aspects of the airline industry such as investment in capacity and merger behavior (cf. Snider (2010), Benkard, Bodoh-Creed, and Lazarev (2010), Benkard (2004)).[12]

A recent paper of note is Berry and Jia (2010). They construct and estimate a discrete choice economic model of the airline industry based on the "BLP" model of Berry, Levinsohn and Pakes (1995). Their goal is to document the reasons for falling profitability of airlines in the 2000s; for instance, they determine that half of the decline in profitability is due to changes in consumer demand. In this chapter we abstract from the supply side of the industry and focus on demand but adopt the instruments they use to correct for the endogeneity of prices.

Another recent paper by Park, Sickles and Simar (2007) - hereafter PSS - estimated price

[12]See also, Reiss and Spiller (1989), Berry (1990, 1992), Ciliberto and Tamer (2009).

elasticities in a semiparametric model of demand in the airline industry. The focus of that paper is the construction of efficient estimators for random effect panel model when parametric assumptions on the random effect are relaxed. PSS show that their estimator achieves the semiparametric lower bound and apply their work to estimate demand for airline travel. We wish to build on their results using more current data and, more importantly, by relaxing the linearity of demand as the source of identification. In other words, we are not restricting the estimated elasticities to be constant at every level of price. The aim of this chapter is different from that of PSS; while they are interested in the efficiency of their estimator, we are more concerned about the potential bias that arises from reliance on the linear functional form.

As described in chapter 1, linear functional forms have advantages in estimation at the expense of misspecification of panel data models. Linearity affords the ability to use random or fixed effects because of the ability to difference between time periods and use the 'within' variance to estimate parameters. Since we cannot subtract equations from each other to eliminate these additive effects, these methods are not available to us.

3 Estimation Models

An econometric model of dynamic demand using panel data is, in its simplest form:

$$Q_{it} = \gamma Q_{it-1} + \beta P_{it} + \alpha_i + u_{it}$$

Interpreting these as structural demand equations, demand in the current period is influenced by demand in the previous period, the current price, and unobservable effects, both permanent (α_i) and time-specific (u_{it}). The data we use in this paper looks at demand by market, so α_i represents market-specific factors unobserved to the econometrician, such as facility improvements at origin or destination airports, among others. Though not written here, we could also include other variables, such as population or other observable variables that describe economic conditions in the origin and destination cities.

Prior studies estimating demand elasticities would estimate these models with either linear IV-type estimators (cf. Hausman-Taylor (1981), Anderson and Hsiao (1991,1992), Arellano and Bond (1991)) or GMM-type estimation (cf. Blundell and Bond (1998), Arellano and Bover (1995)) which either use lags of exogenous variables (IV) or appropriate lags of the dependent variable (GMM) as the exogenous variation used for identification and estimation. Both methods take advantage of the linear form through 'differencing' the equations to remove α_i, the permanent unobserved effect.

PSS expand upon these models to estimate random effects models while assuming that the distributions of the random effects and the initial conditions (the distribution describing (x_{i0}, y_{i0}) can only be described nonparametrically. Specifically, the (key) assumptions that PSS make are:

- α_i is independent of the regressors and is distributed nonparametrically
- (X_i, Y_{i0}) are i.i.d. and are distributed nonparametrically

- The 'within errors' u_{it} are normally distributed

- The model is stationary

They then estimate the following equation:

$$Q_{it} = \gamma Q_{i,t-1} + X_{it}\beta + \alpha_i + \varepsilon_{it}$$

where Y_{it} is revenue passenger mile, X_{it} contains $\ln\left(\frac{own\ price}{mile}\right)$, $\ln\left(\frac{competitors'\ price}{mile}\right)$, and $\ln(\textit{population in origin city})$, and α_i is a route-specific unobserved (random) effect.

When Y_{it} is $\log(Q_{it})$ and the first element of the vector X_{it} is $\log(P_{it})$, β_1 represents the (constant) own-price elasticity of demand where the unit of measurement is the market. When an airline increases its price on any given route by 1%, it can expect a $\beta \times 100$ % increase (typically we expect $\beta < 0$) in its demand. The vector X_{it} also contains the passenger-weighted average of competitors' prices. Its coefficient is the competitor price elasticity of demand - when a competing airline increases its price by 1%, the corresponding increase in demand is $\beta_2 \times 100$. Of course, these interpretations only hold in the absence of responses by other airlines to changes in prices - incorporating those potentially important effects requires a richer model, which we leave for future research.

Continuing with the theme of this dissertation, in this chapter we apply recently developed methods in nonparametric econometrics to models of demand for airline travel. We eliminate the linearity assumption which, in previous papers, is the source of identification for (constant) elasticities of demand. This line of research could also lead to testing procedures for some of the assumptions typically used for estimating demand functions. Instead of full system estimation, we present results from limited-information estimation using instruments in place of prices.

3.1 Nonparametric Estimation

A nonparametric characterization of the above models would be

$$Q_{it} = m(Q_{i,t-1}, P_{it}, X_{it}, \varepsilon_{it})$$

where m denotes an unknown function. This model differs from the models studied in the previous two chapters due to its single-equation structure. Additionally, the assumption that X_{it} and ε_{it} share the same derivative is not made - instead we explicitly make an assumption about the scale/variance of ε_{it} by normalizing the distribution of ε_{it} at a particular value of the vector X_{it}.

The estimated own-price elasticities are not constant and are measured by $\frac{\partial \log Q_{it}}{\partial \log P_{it}} = \frac{\partial m(p_{it})}{\partial p_{it}}$, where m will be various functions that are primarily only restricted to be continuous and monotonic. The techniques we use will estimate the slope of this function with respect to m which we roughly translate to elasticities.

Our estimator uses the method described by Matzkin (2003), which gives us both estimates of the entire function and the derivatives of the function over the support of the exogenous variables. In the results we depict some of these functions to show that the data contain potentially much richer interactions between price and quantity than linear demand models would allow. Due to the endogeneity of prices, however, we do not use $P_{i,t}$ directly and instead use proxies for price in the estimation procedure. This method also differs from the previous chapters because we make the restriction that ε_{it} is no longer autocorrelated.

It is important to note that we must also make an assumption that the demand equation is identified - since we are not restricting the demand and supply functions to have linear form, the nonparametric system must be solvable - the m functions must be such that the right hand side variables correspond to unique values of Q_{it}. It is possible to adapt the methods covered in the first two chapters of this dissertation to deal with the endogeneity of prices and of lagged demand (i.e., to specify conditions to deal specifically with limited-information estimation), but we leave this for future research.

4 Data

The data are taken from the Department of Transportation's DB1B database - a 10% sample of all domestic air travel ticketed in a particular quarter. Our data range from Q1 1993 to Q4 2011. To restrict the definition of the market, we look only at round-trip itineraries between two cities (not airports). We also look only at itineraries ticketed and operated by the same carrier (no 'interlining' itineraries). The quantity demanded is measured by 'revenue-passenger-mile,' which we calculate by taking the number of reported passengers (in the sample, not extrapolated) multiplied by the number of miles flown on the route.

One distinction from some papers, like Berry and Jia (2010), is that we do not distinguish itineraries by direction. We have this information but choose not to use it because in most cases it would be impossible to determine the 'true' origin city. For instance a college student traveling between campus and her home might choose either city as the origin. Also, while back-to-back ticketing is prohibited, nothing prohibits the purchase of two tickets that (completely) overlap, which makes travel in opposite directions potentially well-substitutable. For this paper we assume that travel between cities is the product demanded and make no distinction with regard to direction.

There are a number of factors that were not accounted for in the construction of the data. Chiefly among these is the fact that a number of changes to the market structure occurred during the time spanned by the data. In particular, there were several mergers among the largest participants in the market, such as the acquisition of US Airways by America West (though completed as a "reverse takeover," with US Airways retaining its name and brand), the merger of Delta Air Lines and Northwest Airlines, and the recent merger of United Airlines and Continental Airlines. In addition, there was a failed acquisition of US Airways by United in 2010. There are a number of complications to adjusting the data to reflect merger activity - the merger approval process spans multiple quarters and even after approval, it takes time before airlines fully combine operations. For instance, Delta Air Lines and Northwest Airlines continued to operate under separate operating certificates until January 1, 2010, though the merger was approved by both companies in September

2008 and the Department of Justice approved the merger in October 2008. During Q4 2008 - Q4 2009, the two companies were essentially the same company, yet they continued to sell tickets and operate flights under their distinctive brands (indeed the DB1B data records a number of itineraries for Northwest Airlines even in Q1 of 2010).

Table 1: Summary Statistics

Time period	1993Q1 - 2011Q4
Total Observations:	688,883
Airlines	77
Market-airline pairs	31,175

Occasionally we will use only a subset of these data, particularly when we look at individual airlines. As in Berry and Jia, we describe our data in cross section over a few years. Interestingly our subset of the data (including only round-trip, nondirectional itineraries) paints a slightly different picture from that of Berry and Jia - our data has lower but rising prices over this period. We document the same transition to direct flights - the average number of segments decreases, largely because of more direct flights offered by legacy airlines, even while the average itinerary increased in length.

Table 2: Fare Overview

	1993-2011	1999	2006	2011
Average fare	$372.89	$369.42	$382.61	$440.02
	(136.11)	(141.78)	(131.69)	(156.56)
Average fare (LCC)	$318.56	$272.51	$327.41	$374.71
	(84.55)	(70.82)	(79.78)	(89.30)
Average fare (Legacy)	$388.50	$392.01	$394.67	$473.84
	(140.28)	(146.69)	(134.60)	(171.23)
Trip Descriptors:				
Segments	3.668	3.704	3.535	3.406
	(0.813)	(0.779)	(0.843)	(0.906)
Segments (LCC)	3.414	3.402	3.452	3.410
	(0.840)	(0.846)	(0.830)	(0.843)
Segments (Legacy)	3.723	3.763	3.573	3.410
	(0.799)	(0.761)	(0.838)	(0.935)
Miles	3099.80	3132.23	3230.99	3228.72
	(1728.73)	(1705.35)	(1860.78)	(1857.31)
Miles (LCC)	2839.86	2578.93	2914.07	2926.06
	(1321.99)	(1220.45)	(1364.24)	(1330.71)
Miles (Legacy)	3123.45	3181.11	3230.81	3378.28
	(1788.51)	(1760.36)	(1926.11)	(2064.70)

Our restrictions on the data seem to have led us to include fewer short-haul flights, which might have excluded more business travelers (seeing as our itineraries are longer and cheaper than those of Berry and Jia). The average fares are weighted by number of passengers when calculated. 'Low cost carrier' (LCC) airlines for this table are Airtran, Frontier, Jetblue and Southwest. The 'legacy' airlines are American, Continental, Delta, Northwest, United and US Airways.

Table 3: By Airline

Airline	t	Markets	Total obs	Ave Fare (std dev)
American (AA)	1993Q1 - 2011Q4	3753	96615	397.08 (141.25)
Continental (CO)	1993Q1 - 2011Q4	2861	64135	364.52 (127.28)
Delta (DL)	1993Q1 - 2011Q4	4230	111153	399.55 (141.23)
Northwest (NW)	1993Q1 - 2010Q1	3508	86104	362.51 (124.24)
United (UA)	1993Q1 - 2011Q4	3050	82767	417.14 (156.15)
US Airways (US)	1993Q1 - 2011Q4	2972	68563	378.99 (137.82)
JetBlue (B6)	2002Q2 - 2011Q4	369	5736	345.00 (106.15)
Frontier (F9)	1994Q4 - 2011Q4	742	13420	330.88 (74.53)
AirTran (FL)	1995Q1 - 2011Q4	995	20188	284.83 (63.74)
Southwest (WN)	1998Q3 - 2011Q4	1389	46649	326.91 (85.01)

Broken down by airline, the differences between legacy airlines (toward the top) and LCCs is readily apparent - LCCs are relatively newer, operate in far fewer markets, and have lower prices that are much less dispersed. Of course, much of the fare dispersion is due to one-class service (no first class) in the LCCs (AirTran added "business class" but fare premiums - and extra amenities - are minimal).

4.1 Instruments

Just like in Berry and Jia (2010), we construct our instruments at the route-level and reflect decisions made by competitors on the same origin-destination city pair. The argument for their validity is that decisions by competitors are unrelated to the unobserved drivers of demand for an individual airline but are related to demand through the prices that the airline offers. We use as instruments the percentage of non-stop passengers that are carried by the airline's competitors. We have two measures of our instrument (rivalpass) - since we are looking at round-trip itineraries only, we distinguish between those that have both legs of the trip flown on nonstop flights (rivalpass-both) with those that have at least one leg with a nonstop flight (rivalpass-oneway).[13]

Table 4: Instruments

	Mean	Std. Dev
rivalpass-both	52.86%	(47.39%)
rivalpass-oneway	68.86%	(41.61%)

In order to assess their appropriateness as instruments, we conduct a simple 'first stage' regression in which we (linearly) regress log price per mile on each of our measures of the instrument. We report the coefficient value and the t statistic for the null hypothesis that the coefficient is equal to zero. As expected, the relationship between price and the percent-

[13]The reason for this: if you search delta.com for "nonstop flights only," the results will give you itinerary options with multiple stops on the return leg of the trip.

age of nonstop passengers carried by competitors is negative: the higher the desirability of competition, the more likely prices should decrease. Interestingly, once we convert our price into log(price per mile), the relationship between our instruments and price is positive.

In order to measure price elasticities, then, we will use an 'indirect-least-squares'-type approach to measure the actual price elasticity. This strategy is originally credited to Haavelmo (1943) and most recently was extended to nonseparable/nonparametric models by Schennach, White and Chalak (2009). The method will be informally applied here in the manner, as Schennach, White and Chalak write, of Heckman and Vytlacil's (1999,2001) "local instrumental variable" - the simple local ratio between the derivative of quantity with respect to the proxy and the derivative of price with respect to the proxy; please refer to the paper by Schennach, White and Chalak (2009) for the formal justification.

Table 5: First Stage

Linear - price and instrument

	Coefficient	t-statistic
rivalpass-both	-115.49	-54.07
rivalpass-oneway	-120.57	-56.79

Linear - log(price per mile) and instrument

	Coefficient	t-statistic
rivalpass-both	0.1943	48.36
rivalpass-oneway	0.1387	34.49

To this end we run a nonparametric 'first stage' with own-price as the dependent variable and lagged quantity, competitors' fare (the two exogenous variables in the demand function specification) and the instrument as the 'right-hand-side' variables:

Table 6: First Stage

Nonparametric - log(price per mile) and instrument

	Coefficient
rivalpass-both	-0.0041
rivalpass-oneway	0.0634

Not only is the sign different from expected, the 'second-stage' specifications with $rivalpass - both$ used as a proxy for (own) price were less reliable (especially when estimating by airline separately). This may be because of the greater number of city-pairs with few nonstop options. As a result, unless noted, the reported tables are for specifications where $rivalpass - oneway$ is the proxy variable used.

5 Results

The airline industry itself is, for strategic reasons, interested in measuring the price elasticity of the demand for air travel. The International Air Transport Association (IATA), a trade group, publishes regular economic briefings. One recent briefing focused specifically on research on elasticities for air travel (IATA (2008)). When focusing on the "route-level" as we do (as in, the appropriate elasticity is raising the prices observed on a route in a given time period by 1%) the elasticities measured for North America were in the range of -1.2 to -1.5. PSS report elasticities (for specific airlines) in the range of -2.0 to 0.4.

Table 7: Reported Elasticities (PSS 2005)

Airline	Own Price	Competitor Price
American	-0.4209	-0.0345
Continental	-2.0407	0.2299
Delta	-0.3686	-0.1096
Northwest	-0.2471	-0.2809
United	-0.4475	-0.0566
US Air	-0.3937	-0.1912

5.1 Linear Results

For comparison purposes, we report linear estimates are from a Arellano-Bond (1991) estimator that uses differenced values of the exogenous variables as instruments for the differenced lagged dependent variables.

Table 8: Full Sample Results (Linear)

Coefficient on $Q_{i,t-1}$:	.3879
Own-price elasticity:	-.5663
Cross-price elasticity:	.2757

For the full sample, the autocorrelation of quantity demanded (measured by revenue passenger miles) is around .39. As we expect, the own-price elasticity is negative, but is smaller in absolute value than is reported in many studies. Divided by airline, the results are as we would expect - the legacy airlines have larger price dispersion and face more elastic demand.

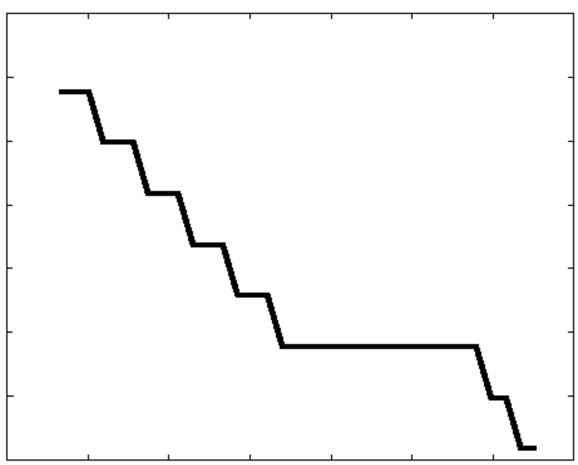
Table 9: By Airline (Linear)

Airline	$Q_{i,t-1}$	Own fare	Comp. fare
DL	.7038	-.8142	.7187
AA	.6083	-.8323	.5436
CO	.5677	-.8195	.4286
UA	.7190	-.4626	.1833
US	.7736	-.5099	.6980
WN	.4095	-.1520	.4842

5.2 Nonparametric Results

The difference between nonparametric and linear demand functions can be depicted in the following graphs

Figure 1: Nonparametric $m(\bar{Q}_{i,t-1}, P^*_{i,t}, \bar{P}_{-i,t}, \bar{\varepsilon}_{it})$**:**



The above graph depicts log(Q_{it} in relation to log(price per mile), as measured by its proxy (P^*_{it} is competitors' share of nonstop passengers on that route), for given values of

$Q_{i,t-1}$, $P_{-i,t}$ and the unobservable variable ε_{it}. Parts of the function do in fact look linear (the functions were estimated on a grid of 27^4 grid); however, there also is some clear price-dependence on the derivative of quantity with respect to price.

Table 10: Average derivative:

	2008-2011
$E\left[\dfrac{\partial m(Q_{i,t-1}, P_{it}, P_{-i,t}, e)}{\partial Q_{i,t-1}}\right]$	0.8228
$E\left[\dfrac{\partial m(Q_{i,t-1}, P_{it}, P_{-i,t}, e)}{\partial P^*_{it}}\right]$	-0.0148
$E\left[\dfrac{\partial m(Q_{i,t-1}, P_{it}, P_{-i,t}, e)}{\partial P_{-i,t}}\right]$	0.0711

We also have the results from the nonparametric first stage, we we can use to measure the 'actual' average derivative of the demand function m with respect to log(price per mile), corrected for the first stage (where log(price per mile) is the dependent variable and $Q_{i,t-1}$, log(competitor fares per mile) and $rivalpass - oneway$).

Table 11: Average derivative:

	2008 - 2011
$E\left[\dfrac{\partial m(Q_{i,t-1}, P_{it}, P_{-i,t}, e)}{\partial P_{it}}\right]$	-0.5726

Looking at the (uncorrected) results by airline:

Table 12: By Airline

	$E\left[\frac{\partial Q_{it}}{\partial Q_{i,t-1}}\right]$	$E\left[\frac{\partial Q_{it}}{\partial P_{it}}\right]$	$E\left[\frac{\partial Q_{it}}{\partial P_{-i,t}}\right]$
Delta	0.7596	-0.0119	0.0351
American	0.7591	-0.0122	0.1303
Continental	0.8528	-0.0214	0.1318
United	0.8161	-0.0150	0.1021
Northwest (N = 5541)	1.0904	-0.0325	0.1240
US Airways	0.6082	-0.0155	0.0914
Southwest	0.6221	-0.0141	0.1395
JetBlue (N = 2433)	1.1820	-0.0111	-0.0280
Virgin (N = 393)	2.3306	-0.1507	0.3914

The average derivatives are comparable to the linear results, though less in magnitude. The elasticities are likely muted because we are using a proxy for the price instead of using methods where we correct for the endogeneity of prices.

As seen from the pictures depicted above, the added benefit from nonparametric estimation is the richer relationship between the variables - to assess the value added we do two things - an experiment wherein we raise the prices that an airline charges in order to assess the implied change (both in average and across fares) in demand, as well as do a simple forecasting exercise where we compare changes in demand predicted by linear and nonparametric methods.

5.2.1 Policy experiment: Price Change

In this policy experiment we suppose that there is a reason that prices for a particular airline would change by 1% uniformly across its entire route network. Though somewhat implausible, there are a number of situations that we could devise that would result in such a scenario. Firstly, there are have been recent surcharges related to airline travel that have been instituted on all airline fares.[14] The most well-known of these might be the 9-11 related surcharge that adds $2.50 to each departure, but there are a number of charges like the excise tax of 7.5%, fees per segment that can be added by airports or the government, and most recently, fuel surcharges that carriers can add to offset the higher price of jet fuel. Carriers are, as of January 2012, required to include all taxes and surcharges in their marketing of ticket prices. Therefore, making the somewhat extreme assumption that there is constant pass-through of a change in the excise tax would mean that prices would change by a fixed percentage across the route network. Besanko, Dube and Gupta (2005) study changes and pass through in retail environments and would warn that this experiment is restricted to "short-run own-brand pass through" of taxes, since we do not account for changes in competitors' prices or possible changes in market equilibria (entry, exit, etc.) due to the change in taxes or prices.

We choose United Airlines as our test case, since its linear estimates and the average derivatives are the closest in size. Supposing that the government, to spur travel, cuts taxes on the lower half of a given carrier's airfares. Suppose the result is a 1% decrease in the price of all tickets below United's median; the linear result is that demand on those flights goes up

[14] Source: http://www.delta.com/planning_reservations/plan_flight/online_reservations/fares_ticketing_rules/taxes_fees/index.jsp

by 0.4626%.[15] The overall result is an increase in demand of 0.2456% in revenue passenger miles. Using our nonparametric estimates, a 1% decrease in the prices of all tickets below United's median results in an increase of of total demand of 2.401%.

5.2.2 Forecasting

As another illustration, we conduct a very informal forecasting exercise where we use our estimated functions to forecast demand out-of-sample. Our nonparametric estimates are from the sample 2008Q1 - 2011Q4, and so we use our results to forecast 2012Q1's demand.

Table 13: Forecasting Results

		Bias (rpm)	RMSE (rpm)
Delta	Linear	-17202	43759
	Nonparametric	11485	29550
United	Linear	-19055	47434
	Nonparametric	10545	31677
Southwest	Linear	-15837	41848
	Nonparametric	19697	33538

The out-of-sample forecasting ability of the nonparametric estimator performs very well compared to the linear estimator, most likely because of the changing and also because of the interaction effects between own-price and competitors' price that are built-in to the non-parametric estimation. More extensive research on the forecasting ability of nonparametric

[15] The median fare for United was $420.97, while the median fare per mile was $0.142.

estimation of dynamic demand seems to deserve attention.

6 Conclusion

The purpose of this paper is to estimate non-constant elasticities over the entire range of airline ticket prices as a robustness check against the possible misspecification of linear models. The benefit to this approach is that in all but a handful of cases, the structural models that are used to describe the individual consumer's choice will lead to route-level demand equations that are not linear. We also apply new results from the field of nonparametric econometrics to give more definitive answers about the identification and estimation of systems of demand and supply equations in the airline industry.

There are a number of avenues for future research. First, the full benefit of the research presented in chapters 1 and 2 are evident if data on the supply side are collected an analyzed. Industry data on costs and supply conditions would make system estimation possible. On a related note, the bridge between structural models and market-level models is a very interesting area for future research. Structural models of individual choice have inputs like conditional choice probabilities that are not able to be estimated consistently under commonly made assumptions (Aguirregabiria and Ho (2011)). While the models in this paper are themselves not necessarily appropriate as inputs into models of discrete choice, the ability to estimate route-level structural supply and demand functions may prove to be useful. In general we would like to pursue this type of research - underlining the link between models of individual choice and how it aggregates into market-level demand (cf. Berry and Haile

(2010)).

The links between structural models of choice and market supply and demand do not exist only on the demand side. Our data span a period with frequent episodes of bankruptcy and mergers, both of which affect pricing and dynamic strategy. Certainly capacity and the supply of seats (which are potentially exactly observable) are correlated with the unobservable (random) route effects. The increased availability of data will allow us to make more definitive statements about these relationships.

This dissertation is focused on promoting the use of nonparametric methods for estimation in situations where endogenous regressor variables would normally force practitioners to use linear methods. Chapter 1 presents basic assumptions to check identification in models with simultaneous equations, nonseparable structural equations and autocorrelated unobservable variables. Chapter 2 describes how to adapt nonparametric estimators of simulataneous equations models for use with the models described in chapter 1. Chapter 3 then relates these models to those that describe supply and demand in the airline industry and performs nonparametric estimation of demand equations (without supply). Further work remains to generalize these models or at least to codify exactly what kinds of 'restrictions' to structural equations minimize and balance the risk associated with model misspecification with slow rates of convergence brought on by the inherent curse of dimensionality in nonparametric estimation.

7 Data Details

The data include itineraries that begin and end in US territories that are outside of the 50 states. Most notably we had to drop tickets that originated or ended in the Virgin Islands, Puerto Rico and Guam. Though technically US territories, we made the decision that these destinations are typically thought of as international destinations and the type of consumer deciding to take visits to Guam would be ones more interested in international destinations.

A few more things about the data - we removed roundtrips where there were reported stops in the same city. We dropped itineraries in which either the outgoing or incoming leg had more than 6 'coupons' (5 stops). We removed 'interline' itineraries (where either the ticketing or operating carriers changed during the itinerary). An important drop was the removal of tickets where the reported fare is less than 10 dollars (likely to be buddy pass/employee/comp/FF tickets) since our focus is on paid travel. We do not make any adjustments for airline quality (both in either the 'hard product' - airframes, physical amenities like seats or entertainment, or the 'soft product' - service/food/other amenities); one huge component of airline quality is the frequent flier program. We also do not include information on the class of service (e.g., first class vs coach). We may be understating the actual price elasticity once we take loyalty into account, however these types of questions are left for future research.

'Bulk fares' were dropped; these fares only make up around 2 itineraries per quarter For those familiar with the literature on the airline industry, in general we decided to make restrictions on the ticket sample that are very similar to the steps taken by Aguirregabiria and Ho (2010).

References

[1] Aguirregabiria, V. and C. Y. Ho (2011). "A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments." *Working Paper.*

[2] Anderson, T. W. and C. Hsiao (1991). "Estimation of Dynamic Models with Error Components." *Journal of the American Statistical Association.*, Vol. 76, No. 375.

[3] Anderson, T. W. and C. Hsiao (1992). "Formulation and Estimation of Dynamic Models using Panel Data." *Journal of Econometrics*, Vol. 18.

[4] 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*, Vol. 58, No. 2.

[5] Arellano, M. and O. Bover (1995). "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, Vol. 68, Issue 1.

[6] Benkard, C. L. (2004). "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft." *The Review of Economic Studies*, Vol. 71.

[7] Benkard, C. L., Bodoh-Creed, A. and J. Lazarev (2010). "Simulating the Dynamic Effects of Horizontal Mergers: U.S. Airlines." *Yale University Working Paper.*

[8] Berry, S. T. (1990). "Airport Presence as Product Differentiation." *American Economic Review*, Vol. 80, No. 2.

[9] Berry, S. T. (1992). "Estimation of a Model of Entry in the Airline Industry." *Econometrica*, Vol. 60, No. 4.

[10] Berry, S. and P. Haile (2009). "Identification of Discrete Choice Demand from Market Level Data." *Yale University Working Paper.*

[11] Berry, S. and P. Jia (2010). "Tracing the Woes: An Empirical Analysis of the Airline Industry." *American Economic Journal: Microeconomics*, Vol 2.

[12] Besanko, D., Dube, J.P. and S. Gupta (2005). "Own-Brand and Cross-Brand retail Pass-Through." *Marketing Science*, Vol. 24, No. 1.

[13] Blundell, R. and S. Bond (1998). "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, Vol. 87.

[14] Blundell, R., Horowitz, J. L. and M. Parey (2011). "Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation." *Quantitative Economics*, Vol. 3.

[15] Choi, S. C. (1991). "Price Competition in a Channel Structure with a Common Retailer." *Marketing Science*, Vol. 10, No. 4.

[16] Ciliberto, F. and E. Tamer (2009). "Market Structure and Multiple Equilibria in Airline Markets." *Econometrica*, Vol. 77, No. 6.

[17] Jaffe, S. and E. G. Weyl (2010). "Linear Demand Systems are Inconsistent with Discrete Choice." *SSRN Working Paper*.

[18] Hausman, J. A. and W. E. Taylor (1981). "Panel Data and Unobservable Individual Effects." *Econometrica*, Vol. 49, No. 6.

[19] IATA. "IATA Economics Briefing: Air Travel Demand." *IATA*, April 2008.

[20] Kelejian, H. (1971). "Two-state least squares and econometric systems linear in parameters but nonliner in endogenous variables." *Journal of the American Statistical Association*, Vol. 66.

[21] Matzkin, R. L. (2003). "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica*, Vol. 71, No. 5.

[22] McGuire, T. W. and R. Staelin (1983). "An Industry Equilibrium Analysis of Downstream Vertical Integration." *Marketing Science*, Vol. 2, No. 2.

[23] Oum, T. H., Waters, W.G. and J. S. Yong (1990). "A Survey of Recent Estimates of Price Elasticities of Demand for Transport." *World Bank Working Paper*.

[24] Park, B. U., Sickles, R. C., and L. Simar (2007). "Semiparametric efficient estimation of dynamic panel data models." *Journal of Econometrics.*, Vol. 136.

[25] Reiss, P. C. and P. T. Spiller (1989). "Competition and Entry in Small Airline Markets." *The Journal of Law and Economics*, Vol. 32, No. 2.

[26] Roehrig, C. (1988). "Conditions for Identification in Nonparametric and Parametric Models." *Econometrica*, Vol. 56, No. 2.

[27] Schennach, S., White, H., and K. Chalak (2009). "Local Indirect Least Squares and Average Marginal Effects in Nonseparable Structural Systems." *UCSD Working Paper*.

[28] Schmalansee, R. and T. Stoker (1999). "Household Gasoline Demand in the United States." *Econometrica*, Vol. 67, No. 3.

[29] Snider, C. (2010). "Predatory Incentives and Predation Policy: The American Airline Case." *Working Paper.*

[30] Su, L., and A. Ullah (2010). "Nonparametric and Semiparametric Panel Econometric Models: Estimation and Testing." *Singapore Management University Working Paper.*