Consistency and Asymptotic Normality for Equilibrium Models with Partially Observed Outcome Variables*

Nathan H. Miller† Georgetown University

Matthew Osborne‡ University of Toronto

November 25, 2013

Abstract

We derive conditions under which structural econometric models that rely on numerical computation of equilibria produce consistent and asymptotically normal parameter estimates. We show that continuity of the equations defining equilibrium, uniqueness of equilibrium, identification of the parameters and two regularity conditions on the first order conditions are sufficient to guarantee that estimates are consistent and asymptotically normal.

Keywords: equilibrium models; consistency; asymptotic normality
JEL classification: C51; L10

*This paper contains results circulated previously in a working paper titled "Competition among Spatially Differentiated Firms: An Estimator with an Application to Cement." We thank Scott Kominers and seminar participants at the Bureau of Economic Analysis, the Federal Trade Commission, George Washington University, Georgetown University, the University of British Columbia – Sauder School of Business, University of Virginia, and the U.S. Department of Justice for valuable comments.

†Corresponding author. Georgetown University, McDonough School of Business, 37th and O Streets NW, Washington DC 20057. (202) 687-0953. Email: nhm27@georgetown.edu

‡University of Toronto, Rotman School of Management, 105 St. George St., Toronto, ON, Canada M5S 3E6. Email: matthew.osborne@rotman.utoronto.ca

1 Introduction

Econometric models that require the computation of equilibria within an estimation routine recently have been introduced in industrial organization. Goettler and Gordon (2011) estimate a dynamic innovation game played between Intel and AMD, and Miller and Osborne (2013) estimate a static game of spatial price discrimination in the cement industry exploiting variation in aggregated endogenous data. Both papers put an inner loop, outer loop structure on estimation. In the inner loop, equilibrium objects such as price or quantity are computed using numerical methods, conditional on a candidate parameter vector, and aggregated to the level of the data. In the outer loop, an optimization algorithm selects the parameter vector that brings the predicted moments closest to the observed moments. The methodology is promising because it relaxes the data requirements of estimation and extends the reach of researchers engaged in structural modeling.

We derive conditions on the equations defining the equilibrium objects that guarantee consistent and asymptotically normal estimates. Standard asymptotics require that the objective function (i) be differentiable at the true parameter value for almost all values of the exogenous variables, and (ii) satisfy a Lipschitz condition in a neighborhood of the true parameter value. While these conditions can be verified if analytical solutions to the equilibrium objects are available (e.g., Thomadsen (2005)), more often such solutions are unavailable and the objects are defined implicitly as solutions to first order conditions.

We first show the equilibrium objects are continuous in the parameters and exogenous variables if a unique equilibrium exists and if the first order conditions are continuous. We then derive two additional conditions that together guarantee consistency and asymptotic normality. First, if the Jacobian of the first order conditions is singular at some vector of exogenous variables, $\mathbf{X}_0$, then a perturbation to that $\mathbf{X}_0$ yields nonsingularity. Second, when the partial derivatives of the equilibrium objects exist they are bounded by a measurable function. Of course, if the first order conditions are always nonsingular, the Implicit Function Theorem (IFT) along with boundedness guarantees asymptotic normality. However, continuity and differentiability of the first order conditions do not guarantee one can apply the IFT. Our conditions thus are weaker than simply assuming the IFT can be used.

2 Model and Estimator

We consider a market that equilibrates in each of $t = 1, \ldots, T$ periods given an exogenous matrix of data $\boldsymbol{X}_t$ and a parameter vector $\boldsymbol{\theta}_0$. Each firm i chooses a strategy vector $\boldsymbol{\sigma}_{it}$

in each period. We assume that each combination of strategy vectors $\boldsymbol{\sigma}_t = (\boldsymbol{\sigma}_{1t}, \boldsymbol{\sigma}_{2t}, \dots)$ produces different equilibrium outcomes. Let the derivative of firm i's profit function with respect to its strategy vector be $\boldsymbol{f}_{it}(\boldsymbol{\sigma}_{it}; \boldsymbol{\sigma}_{-it}, \boldsymbol{X}_t, \boldsymbol{\theta}_0)$. The stacked first order conditions that characterize equilibrium in period t then are given by

$$\boldsymbol{f}_t(\boldsymbol{\sigma}_t; \boldsymbol{X}_t, \boldsymbol{\theta}_0) = \mathbf{0}. \tag{1}$$

We assume that the underlying demand and marginal cost functions are continuous and differentiable in the strategy and parameter vectors, which holds for most standard empirical models. It follows that $\boldsymbol{f}_t(\boldsymbol{\sigma}_{\cdot t}; \boldsymbol{X}_t, \boldsymbol{\theta}_0)$ also has those properties.

Suppose the econometrician observes aggregated equilibrium data such as average prices or total revenue, as well as the exogenous data. Define a continuously differentiable function $S : \mathbb{R}^N \to \mathbb{R}^L$ that maps strategies into aggregated equilibrium data, where N is the length of stacked strategy vectors and L is the number of aggregated equilibrium data points observed. The data generating process is

$$\boldsymbol{Y}_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0) = S(\boldsymbol{\sigma}^*_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0)) + \boldsymbol{\omega}_t, \tag{2}$$

where $\boldsymbol{\sigma}^*_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0)$ is the equilibrium strategy vector and $\boldsymbol{\omega}_t$ is unobserved measurement error.

We assume that the econometrician knows S and can compute the strategy vector $\boldsymbol{\sigma}^*_t(\boldsymbol{\theta}; \boldsymbol{X}_t)$ that solves equation (1) for any given candidate parameter vector $\boldsymbol{\theta}$. Thus, the econometrician can calculate aggregate equilibrium predictions according to $\widetilde{\boldsymbol{Y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t) = S(\boldsymbol{\sigma}^*_t(\boldsymbol{\theta}; \boldsymbol{X}_t))$. This lends itself to the following minimum distance estimator:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{T} \sum_{t=1}^{T} [\boldsymbol{Y}_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0) - \widetilde{\boldsymbol{Y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)]' \boldsymbol{C}_T^{-1} [\boldsymbol{Y}_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0) - \widetilde{\boldsymbol{Y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)], \tag{3}$$

where $\boldsymbol{\Theta}$ is some compact parameter space and $\boldsymbol{C}_T$ is a positive definite matrix. The estimator is analogous to nonlinear least squares, in which each element of $[\boldsymbol{Y}_t(\boldsymbol{X}_t, \theta_0) - \widetilde{\boldsymbol{Y}}_t(\theta; \boldsymbol{X}_t)]$ summarizes a nonlinear equation and $\boldsymbol{C}_T$ weights across equations. Because the aggregated equilibrium predictions are functions of the implicit solution to first-order conditions, additional conditions must be put on the equations that define the equilibrium objects in order for the standard proofs of consistency and asymptotic normality to apply.

3 Asymptotic Properties of the Estimator

It is useful to write the objective function as

$$\frac{1}{T}\sum_{t=1}^{T} m(\boldsymbol{\theta}, \boldsymbol{Y}_t, \boldsymbol{X}_t) \equiv \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{Y}_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0) - S(\boldsymbol{\sigma}^*_t(\boldsymbol{\theta}; \boldsymbol{X}_t)))'\boldsymbol{W}_t(\boldsymbol{Y}_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0) - S(\boldsymbol{\sigma}^*_t(\boldsymbol{\theta}; \boldsymbol{X}_t))), \tag{4}$$

where $\boldsymbol{W}_t \equiv \boldsymbol{C}_t^{-1}$. We assume that $\boldsymbol{W} = \lim_{t\to\infty} \boldsymbol{W}_t$ exists and is positive definite, and that at least one element of $\boldsymbol{X}_t$ is continuous. We denote the distribution and support of $\boldsymbol{X}_t$ as F_x and U, respectively, and denote the distribution of $\boldsymbol{\omega}$ as F_ω.

Assumption A1 (Global Identification): *The parameter vector $\boldsymbol{\theta}_0$ is globally identified in $\boldsymbol{\Theta}$. Formally, $E[\boldsymbol{Y}_t(\boldsymbol{X}_t, \boldsymbol{\theta}_0) - S(\boldsymbol{\sigma}^*_t(\boldsymbol{\theta}; \boldsymbol{X}_t))|\boldsymbol{X}_t] = \boldsymbol{0} \leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

Assumption A2 (Existence and Uniqueness): *For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{X} \in U$ there exists a vector $\boldsymbol{\sigma}_1$ such that $\boldsymbol{f}(\boldsymbol{\sigma}_1; \Psi, \boldsymbol{\theta}) = \boldsymbol{0}$. Further, $\boldsymbol{f}(\boldsymbol{\sigma}_1; \boldsymbol{X}_t, \boldsymbol{\theta}) = \boldsymbol{f}(\boldsymbol{\sigma}_2; \boldsymbol{X}_t, \boldsymbol{\theta}) = \boldsymbol{0} \leftrightarrow \boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$.*

Identification assumptions such as A1 are standard in empirical industrial organization because the conditions for identification in nonlinear models are difficult to formulate and verify theoretically. With partially observed outcomes, A1 could be violated if aggregation is sufficiently coarse. When the properties of the model do not yield A1 and A2 theoretically then the assumptions can be evaluated numerically as in Miller and Osborne (2011).

Lemma 1 (Continuity): *Under A2, the mapping $\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X})$ is continuous in $\boldsymbol{\theta}$ and $\boldsymbol{X}$.*

The corollary that $S(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}))$ is continuous in $\boldsymbol{\theta}$ and $\boldsymbol{X}$ follows from the properties of the aggregating function $S(\cdot)$. Next, since the Jacobian matrix of the first-order conditions need not be nonsingular over all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we cannot rely on the IFT to guarantee that $\boldsymbol{\sigma}^*(\boldsymbol{\theta_0}, \boldsymbol{X})$ is continuously differentiable in a neighborhood of $(\boldsymbol{\theta}, \boldsymbol{X})$. We proceed with the weaker requirement that the first order conditions are well-behaved, in the sense that if the Jacobian is singular at $(\boldsymbol{\theta}_0, \boldsymbol{X}_0)$ then a perturbation to $\boldsymbol{X}_0 + \epsilon$ yields nonsingularity:

Assumption A3: *Denote the Jacobian matrix of the first order condition $\boldsymbol{f}$ with respect to $\boldsymbol{\sigma}$, at $(\boldsymbol{\theta}, \boldsymbol{X})$, as $J\boldsymbol{f}_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{X})$. Consider the true parameter value $\boldsymbol{\theta}_0$ and the set of points*

$$B(\boldsymbol{\theta}_0) = \{\boldsymbol{X} : J\boldsymbol{f}_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X}) \text{ is singular}\}.$$

For each point $\boldsymbol{X}_0$ in $B(\boldsymbol{\theta_0})$, there exists a neighborhood $N(\boldsymbol{X}_0, \boldsymbol{\theta}_0)$ around $\boldsymbol{X}_0$ such that the Jacobian matrix $J\boldsymbol{f}_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X})$ is nonsingular for all $\boldsymbol{X} \in N(\boldsymbol{X}_0, \boldsymbol{\theta}_0)$ and $\boldsymbol{X} \neq \boldsymbol{X}_0$.

Under A3, if the differentiability of $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_0)$ in $\boldsymbol{\theta}$ fails then the IFT guarantees continuous differentiability at the new equilibrium strategy $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_0+\epsilon)$. Provided that one can apply the IFT in an open ball around $\boldsymbol{X}_0$, and at least one element in $\boldsymbol{X}$ is continuously distributed, the set of points at which the IFT fails occurs with zero probability.

Lemma 2 (Differentiability): *Under A2-A3, $S(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}))$ is differentiable in $\boldsymbol{\theta}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for almost all $\boldsymbol{X}$ in U.*[1]

Finally, consistency and asymptotic normality require the objective function to satisfy a Lipschitz condition. The condition holds provided that partial derivatives of $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ almost always exist and can be bounded (when they exist):

Assumption A4: *The partial derivatives of $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ are bounded by a measurable function $M(\boldsymbol{X})$ at all points $\boldsymbol{X} \in U$ for which $J\boldsymbol{f}_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X})$ is nonsingular. Further, for any point $(\boldsymbol{\theta}_0, \boldsymbol{X}_0)$ at which $J\boldsymbol{f}_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_0), \boldsymbol{\theta}_0, \boldsymbol{X}_0)$ is singular, there exists a neighborhood $B(\boldsymbol{\theta}_0)$ in $\boldsymbol{\theta}$-space in which either:*

(i) *For all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0), \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, the partial derivatives of $\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_0)$ with respect to the elements of $\boldsymbol{\theta}$ exist and are bounded by a measurable function $M(\boldsymbol{X})$.*

(ii) *For all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0)$, the partial derivatives of $\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X})$ with respect to the elements of $\boldsymbol{\theta}$ exist in a neighborhood of $\boldsymbol{X}$ around $\boldsymbol{X}_0$, with $\boldsymbol{X} \neq \boldsymbol{X}_0$. These partial derivatives are bounded by a measurable function $M(\boldsymbol{X}) \leq M < \infty$.*

Lemma 3 (Lipschitz Continuity of m in $\boldsymbol{\theta}$-space): *Under A2-A4, there is a measurable function $\dot{m}(\boldsymbol{Y}, \boldsymbol{X})$ such that $|m(\boldsymbol{\theta}_1, \boldsymbol{Y}, \boldsymbol{X}) - m(\boldsymbol{\theta}_2, \boldsymbol{Y}, \boldsymbol{X})| \leq \dot{m}(\boldsymbol{Y}, \boldsymbol{X})\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$ for every $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in some open neighborhood of $\boldsymbol{\theta}_0$*

Proposition 1 (Consistency and Asymptotic Normality): *Under A1-A5 and certain regularity conditions enumerated in the proof,*

(i) plim $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$

(ii) $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to^d N\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\theta}_0}^{-1}\left[\int_U \int_{\boldsymbol{\omega}} \nabla m(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{X}) \nabla m(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{X})' F_x(\boldsymbol{X}) F_\omega(\boldsymbol{\omega})\right] \boldsymbol{V}_{\boldsymbol{\theta}_0}^{-1}\right)$

where $\boldsymbol{V}_{\boldsymbol{\theta}_0}$ is a symmetric matrix that contains the second derivatives of $m(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{X})$ with respect to $\boldsymbol{\theta}$, evaluated at $\boldsymbol{\theta}_0$.

[1]The modifier "almost all" means that the set of $\boldsymbol{X}$ points for which differentiability fails occurs with measure zero, under the probability measure generated by the probability distribution of $\boldsymbol{X}$.

Under Proposition 1, if the parameters are identified globally (A1) and equilibrium is unique (A2) then consistency and asymptotic normality in estimation are guaranteed under weaker conditions than previously recognized. While often it is difficult to verify A1 and A2 theoretically, numerical methods can provide some evidence. Miller and Osborne (2011) support global identification by simulating artificial data from the model using known parameters, aggregating such that the artificial data resemble the market data, and confirming that estimation based on the artificial data recovers the known parameters. It is precisely when A1 and A2 can be verified theoretically or supported numerically that Proposition 1 has the greatest value.

4 Proofs

Proof of Lemma 1: The proof is by contradiction. We note that by assumption $\boldsymbol{f}$ is continuous for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{\sigma}$ in $\mathbb{R}^N$. We suppress $\boldsymbol{X}$ for notational simplicity. The arguments we apply to $\boldsymbol{\theta}$ apply to $\boldsymbol{X}$ as well. Suppose that $\boldsymbol{\sigma}^*(\boldsymbol{\theta})$ is not continuous at $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$. Then there exists an $\epsilon > 0$ such that for all $\delta > 0$ there exists a $\boldsymbol{\theta}_2$ such that

$$0 < \|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1\| < \delta \quad \text{and} \quad \|\boldsymbol{\sigma}^*(\boldsymbol{\theta}_2) - \boldsymbol{\sigma}^*(\boldsymbol{\theta}_1)\| \geq \epsilon. \tag{5}$$

Uniqueness of the equilibrium price $\boldsymbol{\sigma}^*$ implies that if $\|\boldsymbol{\sigma}^*(\boldsymbol{\theta}_2) - \boldsymbol{\sigma}^*(\boldsymbol{\theta}_1)\| \geq \epsilon > 0$, then

$$\|\boldsymbol{f}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1)\| > b > 0.\text{[2]} \tag{6}$$

Continuity of $\boldsymbol{f}$ in $\boldsymbol{\theta}$ implies that for all $\tilde{\epsilon}$ there exists a $\tilde{\delta} > 0$ such that if $0 < \|\boldsymbol{\theta} - \boldsymbol{\theta}_1\| < \tilde{\delta}$,

$$\|\boldsymbol{f}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) - \boldsymbol{f}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}), \boldsymbol{\theta}_1)\| = \|\boldsymbol{f}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}), \boldsymbol{\theta}_1)\| < \tilde{\epsilon}. \tag{7}$$

A contradiction immediately follows from this if we choose $\tilde{\epsilon} = b$. Our initial assertion would imply that for $\tilde{\delta}(b)$ we could find a $\boldsymbol{\theta}_2(\tilde{\delta}(b))$ where

$$0 < \|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1\| < \tilde{\delta} \quad \text{and} \quad \|\boldsymbol{f}(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1)\| \geq b = \tilde{\epsilon}. \tag{8}$$

Proof of Lemma 2: It is sufficient to show that the equilibrium strategy function $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ is almost everywhere differentiable in $\boldsymbol{\theta}_0$. A3 guarantees that for every $\boldsymbol{X}_0$ in

[2] This is because $\|\boldsymbol{\sigma}^*(\boldsymbol{\theta}) - \boldsymbol{\sigma}^*(\boldsymbol{\theta}_1)\| > \epsilon$ implies that $\boldsymbol{\sigma}^*(\boldsymbol{\theta}) \neq \boldsymbol{\sigma}^*(\boldsymbol{\theta}_1)$. Our definition of $\boldsymbol{\sigma}^*$ and the assumption of a unique equilibrium implies $\boldsymbol{f}(\boldsymbol{\sigma}, \boldsymbol{\theta}_1) = \mathbf{0}$ at $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_1)$, and nowhere else.

$B(\boldsymbol{\theta}_0)$, there is an $\boldsymbol{X}$-neighborhood around $\boldsymbol{X}_0$ where the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{\sigma}$ is nonsingular. The IFT guarantees that $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ is continuously differentiable for the $\boldsymbol{X}$ points in this neighborhood. Because each point of possible nondifferentiability $\boldsymbol{X}_0$ is surrounded by an open neighborhood of differentiable points, and at least one element of $\boldsymbol{X}_0$ has a continuous distribution, under the probability measure for $\boldsymbol{X}$ points of nondifferentiability occur with measure zero.

Proof of Lemma 3: First, consider the points $(\boldsymbol{\theta}_0, \boldsymbol{X})$ at which the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{\sigma}$ is nonsingular. At these points, the IFT guarantees that the implicit solution $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ is continuously differentiable in a $\boldsymbol{\theta}$-neighborhood around $\boldsymbol{\theta}_0$ because $\boldsymbol{f}$ is continuously differentiable in $\boldsymbol{\theta}$. It follows that the partial derivatives of $\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ with respect to $\boldsymbol{\theta}$ exist in this neighborhood, and A4 guarantees that the partial derivatives are bounded by $M(\boldsymbol{X})$. Now we turn to $m(\boldsymbol{\theta}, \boldsymbol{Y}_t, \boldsymbol{X}_t)$. Since the $\boldsymbol{W}_t$ has a finite limit, each of its elements $w_{ij,t}$ can be bounded by $\overline{w}_{ij}$.[3] Define

$$m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{Y}_t, \boldsymbol{X}_t) = (Y_{it} - S_i(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))w_{ij,t}(Y_{jt} - S_j(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t)),$$

noting that $m(\boldsymbol{\theta}, \boldsymbol{Y}_t, \boldsymbol{X}_t) = \sum_{i,j} m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{Y}_t, \boldsymbol{X}_t)$. Consider the partial derivative of $m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{Y}_t, \boldsymbol{X}_t)$ with respect to some θ_k. We know that

$$\begin{aligned}
\frac{\partial m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{Y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \quad = \quad & -w_{ij,t}(Y_{jt} - S_j(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))) \cdot \\
& \left[\sum_{n,l} \frac{\partial S_i(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))}{\partial \sigma_{nl}} \frac{\partial \sigma^*_{nl}(\boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial \theta_k} + \frac{\partial S_i(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))}{\partial \theta_k}\right] \\
& -w_{ij,t}(Y_{it} - S_i(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))) \cdot \\
& \left[\sum_{n,l} \frac{\partial S_j(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))}{\partial p_{nl}} \frac{\partial \sigma^*_{nl}(\boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial \theta_k} + \frac{\partial S_j(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))}{\partial \theta_k}\right].
\end{aligned}$$

Our assumption that S is continuously differentiable in its arguments means that there is some $\boldsymbol{\theta}$ neighborhood around $\boldsymbol{\theta}_0$ where $S(\boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{X}_t)$ and the partial derivatives of $S(\boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{X}_t)$ with respect to the elements of $\boldsymbol{\theta}$ are bounded. Moreover, because $\boldsymbol{\sigma}^*$ is continuous in its arguments, it is also bounded in some neighborhood of $\boldsymbol{\theta}_0$. This means that $S(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}_t))$ and its partial derivatives with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ can be bounded in a neighborhood

[3] By the definition of the limit, for all ϵ there is some T for which $|\lim_{t\to\infty} w_{ij,t} - w_{ij,t}| < \epsilon$ for $t > T$. So for all $t > T$, $w_{ij,t}$ can be bounded. $\max_{t \leq T}\{w_{ij,t}\}$ must also exist and be finite, since there are finitely many $w_{ij,t}$'s prior to T. We have implicitly assumed that all the elements of $\boldsymbol{W}_t$ are finite; violations would make numerical maximization of the objective function impossible for some values of t.

of $\boldsymbol{\theta}_0$. We denote the lower bound on S as $\underline{S}$ and the upper bound on the partial derivatives as $\overline{S}'$. Recalling that A4 guarantees that all the partial derivatives of $\boldsymbol{\sigma}^*$ with respect to θ_k are bounded by $|M(\boldsymbol{X}_t)|$, through repeated applications of the triangle inequality we can put a bound on $\frac{\partial m_{ij,t}(\boldsymbol{\theta},\boldsymbol{Y}_t,\boldsymbol{X}_t)}{\partial \theta_k}$:

$$\begin{aligned}\left|\frac{\partial m_{ij,t}(\boldsymbol{\theta},\boldsymbol{Y}_t,\boldsymbol{X}_t)}{\partial \theta_k}\right| &\leq \left|\overline{w}_{ij}\overline{S}'\right|\left(\sum_{n,l}|M(\boldsymbol{X}_t)|+1\right)\left(|y_{it}-\underline{S}|+|y_{jt}-\underline{S}|\right)\\ &= \dot{m}_{ij}(\boldsymbol{Y}_t,\boldsymbol{X}_t).\end{aligned}$$

Recalling that $\boldsymbol{\theta}$ is K dimensional, we can write:

$$\begin{aligned}m(\boldsymbol{\theta}_1,\boldsymbol{Y}_t,\boldsymbol{X}_t)-m(\boldsymbol{\theta}_2,\boldsymbol{Y}_t,\boldsymbol{X}_t) &= \sum_{k=1}^{K} m(\theta_{11},...,\theta_{1k},\theta_{2,k+1},...,\tilde{\theta}_{2K},\boldsymbol{Y}_t,\boldsymbol{X}_t)\\ &\quad -m(\theta_{11},...\theta_{1,k-1},\theta_{2k},...,\theta_{2K},\boldsymbol{Y}_t,\boldsymbol{X}_t)\\ &= \sum_{k=1}^{K}\frac{\partial m(\tilde{\theta}_k,\boldsymbol{Y}_t,\boldsymbol{X}_t)}{\partial\theta_k}(\theta_{1k}-\theta_{2k}).\end{aligned}$$

The second step follows from the Mean Value Theorem (MVT) for $\tilde{\theta}_k=(\theta_{11},...\theta_{1,k-1},\gamma,\theta_{2,k+1},...,\theta_{2K})$, where γ is between θ_{1k} and θ_{2k}. It follows that:

$$\begin{aligned}|m(\boldsymbol{\theta}_1,\boldsymbol{Y}_t,\boldsymbol{X}_t)-m(\boldsymbol{\theta}_2,\boldsymbol{Y}_t,\boldsymbol{X}_t)| &\leq \sum_{k=1}^{K}\left|\frac{\partial m(\tilde{\theta}_k,\boldsymbol{Y}_t,\boldsymbol{X}_t)}{\partial\theta_k}\right||\theta_{1k}-\theta_{2k}|\\ &\leq \sum_{k=1}^{K}\left|\frac{\partial m(\tilde{\theta}_k,\boldsymbol{Y}_t,\boldsymbol{X}_t)}{\partial\theta_k}\right|\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|\\ &\leq K\max_{i,j}\{\dot{m}_{ij}(\boldsymbol{Y}_t,\boldsymbol{X}_t)\}\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|.\end{aligned}$$

Hence, $\dot{m}(\boldsymbol{Y}_t,\boldsymbol{X}_t)=K\max\{\dot{m}_{ij}(\boldsymbol{Y}_t,\boldsymbol{X}_t)\}$ and Lemma 3 holds for the points $(\boldsymbol{\theta}_0,\boldsymbol{X})$ at which the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{\sigma}$ is nonsingular.

Second, we prove the lemma at points of nondifferentiability, i.e., points $(\boldsymbol{\theta}_0,\boldsymbol{X})$ at which the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{\sigma}$ is singular. We first consider Case (ii) of A4 and then return to Case (i). For any $\boldsymbol{X}\neq\boldsymbol{X}_0$, we can argue that

$$\left|\frac{\partial m(\boldsymbol{\theta},\boldsymbol{Y}_t,\boldsymbol{X}_t)}{\partial\theta_k}\right| \leq \left|\overline{w}_{ij}\overline{\overline{S}}'\right|(N|M|+1)\left(\left|y_{it}-\underline{\underline{S}}\right|+\left|y_{jt}-\underline{\underline{S}}\right|\right).$$

This follows from arguments similar to those presented above. A4 (Case (ii)) guarantees that the partial derivatives of $\boldsymbol{\sigma}^*$ are bounded by a constant M. Additionally, since $S(\cdot)$ is continuously differentiable, and since $\boldsymbol{\sigma}^*$ is continuous in our $\boldsymbol{X}$-neighborhood of $\boldsymbol{X}_0$, $S(\cdot)$ and its derivative are bounded by $\underline{\underline{S}}$ and $\overline{\overline{S}}'$, respectively.[4] This implies that the upper bound $\{\dot{m}_{ij}(\boldsymbol{Y}_t, \boldsymbol{X}_t)\}$ is not a function of $\boldsymbol{X}$. It follows that:

$$|m(\boldsymbol{\theta}_1, \boldsymbol{Y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{Y}_t, \boldsymbol{X}_t)| \leq K \max_{i,j}\{\dot{m}_{ij}(\boldsymbol{Y}_t)\}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Taking limits of both sides of this inequality, we see that

$$\begin{aligned}\lim_{\boldsymbol{X}\to\boldsymbol{X}_0} |m(\boldsymbol{\theta}_1, \boldsymbol{Y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{Y}_t, \boldsymbol{X}_t)| &= |m(\boldsymbol{\theta}_1, \boldsymbol{Y}_t, \boldsymbol{X}_0) - m(\boldsymbol{\theta}_2, \boldsymbol{Y}_t, \boldsymbol{X}_0)| \\ &\leq K \max_{i,j}\{\dot{m}_{ij}(\boldsymbol{Y}_t)\}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.\end{aligned}$$

The first line is due to continuity of $\boldsymbol{\sigma}^*$ and $S(\cdot)$. The last line is where the requirement that $M(\boldsymbol{X}_t)$ not be a function of $\boldsymbol{X}_t$ binds. To finish, we turn to Case (i) of A4. We fix $\boldsymbol{X}$ at $\boldsymbol{X}_0$, and again consider applying the mean value theorem to each component of $m(\boldsymbol{\theta}_1, \boldsymbol{Y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{Y}_t, \boldsymbol{X}_t)$. Consider some component

$$m(\theta_{11}, ..., \theta_{1k}, \theta_{2,k+1}, ..., \tilde{\theta}_{2K}, \boldsymbol{Y}_t, \boldsymbol{X}_t) - m(\theta_{11}, ...\theta_{1,k-1}, \theta_{2k}, ..., \theta_{2K}, \boldsymbol{Y}_t, \boldsymbol{X}_t).$$

There are two possibilities to consider. First, suppose that the vector $(\theta_{11}, ...\theta_{1,k-1})$ is different from $(\theta_{01}, ...\theta_{0,k-1})$ in at least one element, or $(\theta_{2,k+1}, ..., \theta_{2K})$ is different from $(\theta_{0,k+1}, ..., \theta_{0K})$ in at least one element. In this case, the vector $\tilde{\theta}_k = (\theta_{11}, ...\theta_{1,k-1}, \gamma, \theta_{2,k+1}, ..., \theta_{2K})$ can never be equal to θ_0. A4 guarantees that the partial derivatives of $\boldsymbol{\sigma}^*$, and hence m, exist for all possible $\tilde{\theta}_k$ so we can apply the single variable MVT as above. The second possibility is that $(\theta_{11}, ...\theta_{1,k-1})$ equals $(\theta_{01}, ...\theta_{0,k-1})$ and $(\theta_{2,k+1}, ..., \theta_{2K})$ equals $(\theta_{0,k+1}, ..., \theta_{0K})$. If $\theta_{1k} = \theta_{2k} = \theta_{0k}$ then the difference above is simply zero. If not, we can prove the following inequality:

$$\frac{|m(\theta_{11}, ..., \theta_{1k}, \theta_{2,k+1}, ..., \tilde{\theta}_{2K}, \boldsymbol{Y}_t, \boldsymbol{X}_t) - m(\theta_{11}, ...\theta_{1,k-1}, \theta_{2k}, ..., \theta_{2K}, \boldsymbol{Y}_t, \boldsymbol{X}_t)|}{|\theta_{1k} - \theta_{2k}|} \leq$$
$$\max\left\{\left|\frac{\partial m(\tilde{\theta}_{1k}, \boldsymbol{Y}_t, \boldsymbol{X}_t)}{\partial \theta_k}\right|, \left|\frac{\partial m(\tilde{\theta}_{2k}, \boldsymbol{Y}_t, \boldsymbol{X}_t)}{\partial \theta_k}\right|\right\}.$$

To prove this, define $g(\gamma) = m(\theta_{11}, ..., \gamma, \theta_{2,k+1}, ..., \tilde{\theta}_{2K}, \boldsymbol{Y}_t, \boldsymbol{X}_t)$. Assuming without loss of generality that $\theta_{1k} < \theta_{2k}$, from A4 we know that $g(\gamma)$ is differentiable on the open intervals

[4] If the $\boldsymbol{X}$ neighborhood is large enough that they are not bounded, we can simply shrink the neighborhood until they are.

$(\theta_{1k}, \theta_{0k})$ and $(\theta_{0k}, \theta_{2k})$ and it is continuous on the interval $[\theta_{1k}, \theta_{2k}]$ due to continuity of S and $\boldsymbol{\sigma}^*$. Hence we can apply the MVT on the interval $(\theta_{1k}, \theta_{0k})$ and $(\theta_{0k}, \theta_{2k})$ to show that

$$\frac{|g(\theta_{0k}) - g(\theta_{1k})|}{|\theta_{0k} - \theta_{1k}|} \leq \left| \frac{\partial m(\tilde{\theta}_{1k}, \boldsymbol{Y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right|, \text{ and } \frac{|g(\theta_{2k}) - g(\theta_{0k})|}{|\theta_{2k} - \theta_{0k}|} \leq \left| \frac{\partial m(\tilde{\theta}_{2k}, \boldsymbol{Y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right|,$$

for some $\tilde{\theta_{1k}} \in (\theta_{1k}, \theta_{0k})$ and $\tilde{\theta_{2k}} \in (\theta_{0k}, \theta_{2k})$. We next show that

$$\frac{|g(\theta_{2k}) - g(\theta_{1k})|}{|\theta_{2k} - \theta_{1k}|} \leq \max \left\{ \frac{|g(\theta_{0k}) - g(\theta_{1k})|}{|\theta_{0k} - \theta_{1k}|}, \frac{|g(\theta_{2k}) - g(\theta_{0k})|}{|\theta_{2k} - \theta_{0k}|} \right\}.$$

To show this inequality, we first make the following definitions:

$$m_1 = \frac{g(\theta_{2k}) - g(\theta_{1k})}{\theta_{2k} - \theta_{1k}}, \quad m_2 = \frac{g(\theta_{0k}) - g(\theta_{1k})}{\theta_{0k} - \theta_{1k}}, \quad m_3 = \frac{g(\theta_{2k}) - g(\theta_{0k})}{\theta_{2k} - \theta_{0k}}.$$

Then define three lines on the interval $[\theta_{1k}, \theta_{2k}]$:

$$L_1(\theta) = m_1\theta + b_1, \quad L_2(\theta) = m_2\theta + b_2, \quad L_3(\theta) = m_3\theta + b_3, \tag{9}$$

where we define

$$b_1 = g(\theta_1) - m_1\theta_1, \quad b_2 = g(\theta_1) - m_2\theta_1, \quad \text{and} \quad b_3 = g(\theta_2) - m_3\theta_2. \tag{10}$$

Because of the way we have defined these lines, and because of the continuity of g, it must be the case that $L_2(\theta_0) = L_3(\theta_0)$, $L_1(\theta_1) = L_2(\theta_1)$, and $L_1(\theta_2) = L_3(\theta_2)$. Let us suppose by way of contradiction that $|m_1| > \max\{|m_2|, |m_3|\}$. There are a number of cases that we have to consider. First, suppose that m_1, m_2, and m_3 are all positive. Then it must be the case that for $\theta > \theta_1$, $L_1(\theta) > L_2(\theta)$ since $L_1(\theta_1) = L_2(\theta_1)$ and L_1 has a steeper slope than L_2. It must also be the case that for $\theta < \theta_2$, $L_1(\theta) < L_3(\theta)$ since L_1 is more steep than L_3 and $L_1(\theta_2) = L_3(\theta_2)$. Since $\theta_1 < \theta_0 < \theta_2$, this implies that $L_3(\theta_0) > L_1(\theta_0) > L_2(\theta_0)$. This contradicts $L_2(\theta_0) = L_3(\theta_0)$. Next suppose that $m_1 > 0$, $m_2 < 0$, and $m_3 > 0$. It is easy to show that it must be the case that $L_2(\theta_0) < L_1(\theta_0)$ (because L_2 slopes down from θ_1, while L_1 slopes upward), and $L_3(\theta_0) > L_1(\theta_0)$ (by the assumption that $m_1 > m_3$), again leading to a contradiction. Then suppose that $m_1 > 0$, $m_2 > 0$ and $m_3 < 0$. The assumption that $m_1 > m_2$ implies that $L_2(\theta_0) < L_1(\theta_0)$. Since we assumed that m_3 is negative, L_3 slopes up from $\theta < \theta_2$ and L_1 slopes down, implying that $L_3(\theta_0) > L_1(\theta_0)$. This again is a contradiction of $L_2(\theta_0) = L_3(\theta_0)$. The cases where $m_1 < 0$ can be shown with similar logic.

The fact that $|m_1| \leq \max\{|m_2|, |m_3|\}$ implies that

$$\frac{|g(\theta_{1k}) - g(\theta_{2k})|}{|\theta_{1k} - \theta_{2k}|} \leq \max\left\{\left|\frac{\partial m(\tilde{\theta}_{1k}, \boldsymbol{Y}_t, \boldsymbol{X}_t)}{\partial \theta_k}\right|, \left|\frac{\partial m(\tilde{\theta}_{2k}, \boldsymbol{Y}_t, \boldsymbol{X}_t)}{\partial \theta_k}\right|\right\}.$$

Similar logic to what was used to prove the last two cases can be used to show

$$|m(\boldsymbol{\theta}_1, \boldsymbol{Y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{Y}_t, \boldsymbol{X}_t)| \quad \leq \quad K \max_{i,j}\{\dot{m}_{ij}(\boldsymbol{Y}_t, \boldsymbol{X}_t)\}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Proof of Proposition 1: With Lemmas 1-3 in hand, the proof of Proposition 1 follows directly from Theorem 5.23 in van der Vaart (1998), pages. 53-54. Two additional normalcy conditions are required:

(i) $E_{\boldsymbol{\omega}} E_{\boldsymbol{X}} \dot{m}(\boldsymbol{Y}, \boldsymbol{X})^2 < \infty$.

(ii) The mapping $\boldsymbol{\theta} \to Pm(\boldsymbol{\theta})$ admits a second-order Taylor expansion at $\boldsymbol{\theta}_0$ such that

$$\begin{aligned} \boldsymbol{\theta} \to Pm(\boldsymbol{\theta}) \quad = \quad & \int_U \left[(S(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X})) - S(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}))' \cdot \right. \\ & \boldsymbol{W}\left(S(\boldsymbol{\sigma}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_t)) - S(\boldsymbol{\sigma}^*(\boldsymbol{\theta}, \boldsymbol{X}))\right)\Big] F_x(\boldsymbol{X}) + E\boldsymbol{\omega}'\boldsymbol{W}\boldsymbol{\omega}. \end{aligned}$$

References

Goettler, R. and B. Gordon (2011). Does AMD spur Intel to innovate more? *Journal of Political Economy 119*(6), 1141–1200.

Miller, N. H. and M. Osborne (2011). Competition among spatially differentiated firms: An estimator with an application to cement. *BEA Working Paper WP2011-04*.

Miller, N. H. and M. Osborne (2013). Spatial differentiation and price discrimination in the cement industry: Evidence from a structural model. *RAND Journal of Economics accepted*.

Thomadsen, R. (2005). The effect of ownership structure on prices in geographically differentiated industries. *RAND Journal of Economics 36*, 908–929.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.