Supplement to "Estimating dynamic discrete-choice games of incomplete information"

(*Quantitative Economics*, Vol. 6, No. 3, November 2015, 567–597)

Michael Egesdal
Department of Economics, Harvard University

Zhenyu Lai
Department of Economics, Harvard University

Che-Lin Su
The University of Chicago Booth School of Business

In Section S1 of this supplement, we provide the replication of the proof for large-sample properties of the maximum-likelihood (ML) estimator.

S1. Large-sample properties of the ML estimator

In this section, we establish the large-sample properties of the ML estimator solved by a constrained optimization approach. Recall that the constrained optimization formulation of the ML estimation problem is

$$\begin{aligned} &\max_{(\boldsymbol{\theta},\mathbf{P},\mathbf{V})} \frac{1}{M}\mathcal{L}(\mathbf{Z};\mathbf{V},\mathbf{P},\boldsymbol{\theta}) \\ &\text{subject to} \quad \mathbf{V}=\Psi^{\mathbf{V}}(\mathbf{V},\mathbf{P},\boldsymbol{\theta}), \\ &\qquad\qquad\quad\; \mathbf{P}=\Psi^{\mathbf{P}}(\mathbf{V},\mathbf{P},\boldsymbol{\theta}). \end{aligned} \tag{S1}$$

We formulate the ML estimation problem (S1) in the framework of Aitchison and Silvey (1958). In Section S1.1, we state the theorem and proves existence, consistency, and asymptotic normality of the ML estimator under a set of conditions analogous to those provided in Aitchison and Silvey (1958). Similar results are also stated in Section 10.3 in Gourieroux and Monfort (1995).

Let $\boldsymbol{\gamma}=(\mathbf{V},\mathbf{P},\boldsymbol{\theta})\in\mathbb{R}^r$ represent the vector containing the choice probabilities, expected value functions, and structural parameters of the dynamic game. A solution of

Michael Egesdal: egesdal@fas.harvard.edu
Zhenyu Lai: zlai@wayfair.com
Che-Lin Su acknowledges the financial support from the Neubauer Family Faculty Fellowship at The University of Chicago Booth School of Business.

DOI: 10.3982/QE430

the dynamic game satisfies the system of equations

$$\begin{aligned} &\mathbf{V} - \Psi^{\mathbf{V}}(\mathbf{V}, \mathbf{P}, \boldsymbol{\theta}) = 0, \\ &\mathbf{P} - \Psi^{\mathbf{P}}(\mathbf{V}, \mathbf{P}, \boldsymbol{\theta}) = 0. \end{aligned} \tag{S2}$$

Let h, a function from $\mathbb{R}^r$ to $\mathbb{R}^s$ with $h(\boldsymbol{\gamma}) = 0$, represent the system of constraint equations (S2) above. The ML estimator solves the constrained optimization problem

$$\begin{aligned} &\max_{\boldsymbol{\gamma}} \frac{1}{M} \mathcal{L}(\mathbf{Z}; \boldsymbol{\gamma}) \\ &\quad \text{subject to} \quad h(\boldsymbol{\gamma}) = 0, \end{aligned} \tag{S3}$$

where $\mathbf{Z}$ denotes observed data and $\mathcal{L}(\mathbf{Z}; \boldsymbol{\gamma})$ is the logarithm of the likelihood function

$$\begin{aligned} \mathcal{L}(\mathbf{Z}; \boldsymbol{\gamma}) &= \log\left(\prod_{m=1}^{M} \prod_{t=1}^{T} \prod_{i=1}^{N} \Psi_i^{\mathbf{P}}\left(\bar{a}_i^{mt} | \bar{\mathbf{x}}^{mt}; \mathbf{V}, \mathbf{P}, \boldsymbol{\theta}\right) \right) \\ &= \sum_{m=1}^{M} \log\left(\prod_{t=1}^{T} \prod_{i=1}^{N} \Psi_i^{\mathbf{P}}\left(\bar{a}_i^{mt} | \bar{\mathbf{x}}^{mt}; \mathbf{V}, \mathbf{P}, \boldsymbol{\theta}\right) \right). \end{aligned}$$

For markets $m = 1, 2, \ldots, M$, let $\mathbf{z}_m$ be the vector of observations with probability density function $f(\mathbf{z}_m, \boldsymbol{\gamma}_0)$, where $\boldsymbol{\gamma}_0$ is the true parameter vector of the data generating process. Here $f(\mathbf{z}_m, \boldsymbol{\gamma})$ is given by $\prod_{t=1}^{T} \prod_{i=1}^{N} P_i(\bar{a}_i^{mt} | \bar{\mathbf{x}}^{mt})$, and the random vectors $\mathbf{z}_m$ are independently and identically distributed across markets. We rewrite the objective function as

$$\mathcal{L}(\mathbf{Z}; \boldsymbol{\gamma}) = \sum_{m=1}^{M} \log f(\mathbf{z}_m, \boldsymbol{\gamma}).$$

Let $\mathbb{L}(\mathbf{Z}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \frac{1}{M}\mathcal{L}(\mathbf{Z}; \boldsymbol{\gamma}) + h(\boldsymbol{\gamma})'\boldsymbol{\lambda}$ be the Lagrangian function, where the vector of Lagrange multipliers is $\boldsymbol{\lambda} \in \mathbb{R}^s$. The ML estimator $\hat{\boldsymbol{\gamma}}$, along with Lagrange multipliers $\hat{\boldsymbol{\lambda}}$, are a solution to the system of equations

$$\begin{aligned} &\frac{1}{M} \nabla_{\boldsymbol{\gamma}} \mathcal{L}(\mathbf{Z}; \boldsymbol{\gamma}) + \nabla_{\boldsymbol{\gamma}} h(\boldsymbol{\gamma})^{\mathrm{T}} \boldsymbol{\lambda} = 0, \\ &h(\boldsymbol{\gamma}) = 0. \end{aligned} \tag{S4}$$

S1.1 *Existence, consistency, and asymptotic normality*

Theorem S1 establishes the large-sample properties of the ML estimator. The proof and its required assumptions are adapted from Aitchison and Silvey (1958). We use $\mathcal{F}$ and $\mathcal{H}$ to label the assumptions required on the likelihood and constraint functions, respectively. The assumptions are as follows.

($\mathcal{F}1$) There is a true $\boldsymbol{\gamma}_0 \in \mathbb{R}^r$, and for all $\boldsymbol{\gamma}$ within an α neighborhood $\mathcal{N}^{\alpha}_{\boldsymbol{\gamma}_0} = \{\boldsymbol{\gamma} : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| \leq \alpha\}$, the following statements hold:

(a) There exist probability density functions $f(\mathbf{z}, \boldsymbol{\gamma})$ with $\mathbf{z} \in \mathbb{R}^q$.

(b) The derivatives $\frac{\partial \log f(\mathbf{z},\boldsymbol{\gamma})}{\partial \gamma_i}$, $\frac{\partial^2 \log f(\mathbf{z},\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}$, and $\frac{\partial^3 \log f(\mathbf{z},\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j \partial \gamma_k}$ exist, for $i, j = 1, \ldots, s$ and $k = 1, \ldots, q$.

(c) The first and second derivatives $\frac{\partial \log f(\mathbf{z},\boldsymbol{\gamma})}{\partial \gamma_i}$ and $\frac{\partial^2 \log f(\mathbf{z},\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}$ are continuous and bounded by finitely integrable functions $F_1(\mathbf{z})$ and $F_2(\mathbf{z})$, and the third derivatives $\frac{\partial^3 \log f(\mathbf{z},\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j \partial \gamma_k}$ are bounded by a function $F_3(\mathbf{z})$ with a finite expectation for $i, j = 1, \ldots, s$ and $k = 1, \ldots, q$.

($\mathcal{F}2$) The information matrix $\mathcal{I}(\boldsymbol{\gamma}_0) = -E[\frac{\partial^2 \log f(\mathbf{z},\boldsymbol{\gamma_0})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}]$ exists and is positive definite with minimum latent root μ_0.

($\mathcal{H}1$) There is a continuous function $h(\boldsymbol{\gamma}) : \mathbb{R}^r \mapsto \mathbb{R}^s$ such that $h(\boldsymbol{\gamma}_0) = 0$, $s < r$ and for all $\boldsymbol{\gamma} \in \mathcal{N}^{\alpha}_{\boldsymbol{\gamma}_0}$, the following statements hold:

(a) The partial derivatives $\frac{\partial h_k(\boldsymbol{\gamma})}{\partial \gamma_i}$ exist and are continuous for $i = 1, \ldots, r$ and $k = 1, \ldots, s$.

(b) The partial derivatives $\frac{\partial^2 h_k(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}$ exist and are bounded for $i, j = 1, \ldots, r$ and $k = 1, \ldots, s$.

(c) The matrix $\nabla_{\boldsymbol{\gamma}} h(\boldsymbol{\gamma}_0)$ is of rank s.

THEOREM S1. *Suppose assumptions $\mathcal{F}1$, $\mathcal{F}2$, and $\mathcal{H}1$ hold. Then the following statements hold*:

(a) Existence. *For an arbitrarily small $\delta > 0$ and any $0 < \varepsilon < 1$, there exists $M_{\varepsilon,\delta}$ such that if $M > M_{\varepsilon,\delta}$, there exists with probability greater than $1 - \varepsilon$ a solution $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}})$ to the constrained maximum-likelihood problem defined by* (S4) *with $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| < \delta$.*

(b) Consistency Under Uniqueness. *If there exists M_0 such that a solution to* (S4) *is unique for all $M > M_0$, then $\hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$.*

(c) Asymptotic Normality. *We have $\sqrt{M}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0, \hat{\boldsymbol{\lambda}}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} \boldsymbol{\Sigma}_{\gamma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \end{smallmatrix}\right)$. Let $\boldsymbol{\mathcal{I}}_0 = \mathcal{I}(\boldsymbol{\gamma}_0)$, let I_r be the $r \times r$ identity matrix, and let $\mathbf{H}_0 = \nabla_{\boldsymbol{\gamma}} h(\boldsymbol{\gamma}_0) \in \mathbb{R}^{s\times r}$. Then $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is the $r \times r$ matrix $\boldsymbol{\mathcal{I}}_0^{-1}(\mathrm{I}_r - \mathbf{H}_0^{\mathrm{T}}(\mathbf{H}_0\boldsymbol{\mathcal{I}}_0^{-1}\mathbf{H}_0^{\mathrm{T}})^{-1}\mathbf{H}_0\boldsymbol{\mathcal{I}}_0^{-1})$ and $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}$ is the $s \times s$ matrix $(\mathbf{H}_0\boldsymbol{\mathcal{I}}_0^{-1}\mathbf{H}_0^{\mathrm{T}})^{-1}$. The rank of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is $r - s$.*

Following the steps in Aitchison and Silvey (1958), we prove existence and consistency under uniqueness in Section S1.1.1, and asymptotic normality in Section S1.1.2.

S1.1.1 *Proof of existence and consistency* Let $\|\cdot\|$ denote the norm operator in Euclidean space. Denote an α neighborhood of $\boldsymbol{\gamma}_0$ using $\mathcal{N}^{\alpha}_{\boldsymbol{\gamma}_0} = \{\boldsymbol{\gamma} : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| \leq \alpha\}$. Let

$\delta < \min\{\alpha, 1\}$, and consider $\boldsymbol{\gamma} \in \mathcal{N}^{\delta}_{\boldsymbol{\gamma}_0}$. First, expand (S4) about $\boldsymbol{\gamma}_0$ and obtain

$$\begin{aligned} &\frac{1}{M}\nabla_{\boldsymbol{\gamma}}\mathcal{L}(\mathbf{Z},\boldsymbol{\gamma}_0) + \frac{1}{M}\frac{\partial^2\mathcal{L}(\mathbf{Z},\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}^2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \frac{1}{M}\tilde{r}_1(\mathbf{Z},\boldsymbol{\gamma}) + \nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}\boldsymbol{\lambda} = 0, \\ &h(\boldsymbol{\gamma}_0) + \nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \tilde{r}_2(\boldsymbol{\gamma}) = 0, \end{aligned} \tag{S5}$$

where $\tilde{r}_1$ and $\tilde{r}_2$ denote remainder terms involving higher order derivatives. From $\mathcal{F}1$ and $\mathcal{H}1$,

(i) $\frac{1}{M}\nabla_{\boldsymbol{\gamma}}\mathcal{L}(\mathbf{Z},\boldsymbol{\gamma}_0) + \frac{1}{M}\tilde{r}_1(\mathbf{Z},\boldsymbol{\gamma}) = o_p(1) + o_p(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|)$,

(ii) $-\frac{1}{M}\frac{\partial^2\mathcal{L}(\mathbf{Z},\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} - \mathcal{I}(\boldsymbol{\gamma}_0) = o_p(1)$,

(iii) $\tilde{r}_2(\boldsymbol{\gamma}) = O(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|^2)$,

(iv) $h(\boldsymbol{\gamma}_0) = 0$.

Substituting (i)–(iv) into system (S5) leads to

$$\begin{aligned} &-\mathcal{I}(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}\boldsymbol{\lambda} + o_p(1) - o_p(1)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + o_p\big(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|\big) = 0, \\ &\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + O\big(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|^2\big) = 0. \end{aligned} \tag{S6}$$

Since $\delta < 1$, we can rewrite equation (S6) as

$$-\mathcal{I}(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}\boldsymbol{\lambda} + o_p(1) = 0, \tag{S7}$$

$$\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + O\big(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|^2\big) = 0. \tag{S8}$$

Assumptions $\mathcal{F}2$ and $\mathcal{H}1$(c) allow us to premultiply (S7) by $\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)\mathcal{I}(\boldsymbol{\gamma}_0)^{-1}$ and obtain the equation

$$-\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)\mathcal{I}(\boldsymbol{\gamma}_0)^{-1}\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}\boldsymbol{\lambda} + o_p(1) = 0. \tag{S9}$$

Assumptions $\mathcal{F}2$ and $\mathcal{H}1$(c) imply that $\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)\mathcal{I}(\boldsymbol{\gamma}_0)^{-1}\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)^{\mathrm{T}}$ is invertible. When δ is sufficiently small, $\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)\mathcal{I}(\boldsymbol{\gamma}_0)^{-1}\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}$ will also be invertible. Inverting (S9) and substituting in equation (S8) yields

$$\begin{aligned} \boldsymbol{\lambda} &= \big[\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)\mathcal{I}(\boldsymbol{\gamma}_0)^{-1}\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}\big]^{-1}\big(\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\big) + o_p(1) \\ &= \big[\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)\mathcal{I}(\boldsymbol{\gamma}_0)^{-1}\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}\big]^{-1}\big(O\big(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|^2\big)\big) + o_p(1) \\ &= O\big(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|^2\big) + o_p(1). \end{aligned} \tag{S10}$$

Substitute (S11) back into (S7). Consolidating the $o_p(1)$ terms, we have

$$-\mathcal{I}(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma})^{\mathrm{T}}O\big(\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|^2\big) + o_p(1) = 0. \tag{S11}$$

Assumption $\mathcal{H}1(b)$ allows us to rewrite (S11) as

$$-\mathcal{I}(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \delta^2 v(\boldsymbol{\gamma}) + o_p(1) = 0, \tag{S12}$$

where $v(\boldsymbol{\gamma})$ is a bounded continuous function of $\boldsymbol{\gamma}$, so that $\|v(\boldsymbol{\gamma})\| < K$.

We will now make use of a result that is equivalent to Brouwer's fixed-point theorem. Interested readers may find a proof of the following lemma in Aitchison and Silvey (1958).

Lemma. *If g is a continuous function mapping $\mathbb{R}^r$ into itself with the property that, for every γ such that $\|\gamma\| = 1$, $\gamma' g(\gamma) < 0$, then there exists a point $\hat{\gamma}$ such that $\|\hat{\gamma}\| < 1$ and $g(\hat{\gamma}) = 0$.*

From (S12), we define a function g on the unit sphere in $\mathbb{R}^r$ as

$$g\left(\frac{\boldsymbol{\gamma} - \boldsymbol{\gamma}_0}{\delta}\right) = -\mathcal{I}(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + \delta^2 v(\boldsymbol{\gamma}) + o_p(1). \tag{S13}$$

Fix ε such that $0 < \varepsilon < 1$. Pick δ small and let $M_{\varepsilon,\delta}$ be such that for any $M > M_{\varepsilon,\delta}$, we have $\Pr(\|o_p(1)\| < \delta^2) > 1 - \varepsilon$. When M is sufficiently large, we then have with probability greater than $1 - \varepsilon$ that

$$\begin{aligned}
&\frac{1}{\delta}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}} g\left(\frac{\boldsymbol{\gamma} - \boldsymbol{\gamma}_0}{\delta}\right) \\
&\quad = -\frac{1}{\delta}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}} I(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + \delta(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}} v(\boldsymbol{\gamma}) + \frac{1}{\delta}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}} o_p(1) \\
&\quad \le -\frac{1}{\delta}\mu_0 \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\|^2 + \delta K \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| + \frac{\delta^2}{\delta}\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\|,
\end{aligned} \tag{S14}$$

where the inequality follows from assumptions $\mathcal{F}2$ and $\mathcal{H}1$(b). Choosing $\boldsymbol{\gamma}$ such that $\delta = \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\|$, we have with probability greater than $1 - \varepsilon$ that

$$\frac{1}{\delta}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}} g\left(\frac{\boldsymbol{\gamma} - \boldsymbol{\gamma}_0}{\delta}\right) \le -\delta\mu_0 + \delta^2 K + \delta^2 < 0 \tag{S15}$$

if δ is sufficiently small and M is sufficiently large.

Applying the lemma, we have that for an arbitrarily small $\delta > 0$, and any $0 < \varepsilon < 1$, there exists $M_{\varepsilon,\delta}$ such that if $M > M_{\varepsilon,\delta}$, there is, with probability greater than $1 - \varepsilon$, a solution to our problem with $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| < \delta$. Then, as long as the solution to problem (S4) is unique when M is sufficiently large, the solution $\hat{\boldsymbol{\gamma}}$ is consistent.

S1.1.2 *Proof of asymptotic normality* Suppose that $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}})$ is a solution to the system of equations (S4). Further suppose that the solution to problem (S4) is unique when M is sufficiently large, so that the solution is consistent. At $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}})$, the system of equations is

$$\begin{aligned}
&\frac{1}{M}\frac{\partial^2 \mathcal{L}(\mathbf{Z}, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}^2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \frac{1}{M}\tilde{r}_1(\mathbf{Z}, \hat{\boldsymbol{\gamma}}) + \nabla_{\boldsymbol{\gamma}} h(\hat{\boldsymbol{\gamma}})^{\mathrm{T}}\hat{\boldsymbol{\lambda}} = -\frac{1}{M}\nabla_{\boldsymbol{\gamma}}\mathcal{L}(\mathbf{Z}, \boldsymbol{\gamma}_0), \\
&\nabla_{\boldsymbol{\gamma}} h(\boldsymbol{\gamma}_0)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \tilde{r}_2(\hat{\boldsymbol{\gamma}}) = \mathbf{0}.
\end{aligned} \tag{S16}$$

We can rewrite (S16) by grouping the remainder terms with the expected value of the derivatives at the true value of $\boldsymbol{\gamma}_0$:

$$-\big(\mathcal{I}(\boldsymbol{\gamma}_0)+\hat{\boldsymbol{I}}\big)(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)+\big(\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)^{\mathrm{T}}+\hat{\boldsymbol{H}}\big)\hat{\boldsymbol{\lambda}}=-\frac{1}{M}\nabla_{\boldsymbol{\gamma}}\mathcal{L}(\mathbf{Z},\boldsymbol{\gamma}_0),$$
$$\big(\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)+\tilde{\boldsymbol{H}}\big)(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)=\mathbf{0}. \tag{S17}$$

Because $\hat{\boldsymbol{\gamma}}$ is consistent, we have $\hat{\boldsymbol{I}}=O(\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|)+o_p(1)=o_p(1)$, $\hat{\boldsymbol{H}}=O(\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|)=o_p(1)$, and $\tilde{\boldsymbol{H}}=O(\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|)=o_p(1)$.

Let $\mathcal{I}_0=\mathcal{I}(\boldsymbol{\gamma}_0)$ and $\mathbf{H}_0=\nabla_{\boldsymbol{\gamma}}h(\boldsymbol{\gamma}_0)$. The linear system (S17) in matrix form is

$$\begin{bmatrix}-(\boldsymbol{\mathcal{I}}_0+\hat{\boldsymbol{I}}) & (\mathbf{H}_0^{\mathrm{T}}+\hat{\boldsymbol{H}})\\ (\mathbf{H}_0+\tilde{\boldsymbol{H}}) & \mathbf{0}\end{bmatrix}\cdot\begin{bmatrix}\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\\ \hat{\boldsymbol{\lambda}}\end{bmatrix}=\begin{bmatrix}-\frac{1}{M}\nabla_{\boldsymbol{\gamma}}\mathcal{L}(\mathbf{Z},\boldsymbol{\gamma}_0)\\ \mathbf{0}\end{bmatrix}. \tag{S18}$$

Assumptions $\mathcal{F}2$ and $\mathcal{H}1$ imply that the matrix $\left[\begin{smallmatrix}-\boldsymbol{\mathcal{I}}_0 & \mathbf{H}_0^{\mathrm{T}}\\ \mathbf{H}_0 & \mathbf{0}\end{smallmatrix}\right]$ is nonsingular. If M is sufficiently large, the matrix $\left[\begin{smallmatrix}-(\boldsymbol{\mathcal{I}}_0+\hat{\boldsymbol{I}}) & (\mathbf{H}_0^{\mathrm{T}}+\hat{\boldsymbol{H}})\\ (\mathbf{H}_0+\tilde{\boldsymbol{H}}) & \mathbf{0}\end{smallmatrix}\right]$ will also be nonsingular with arbitrarily high probability. Since $-\frac{1}{\sqrt{M}}\nabla_{\boldsymbol{\gamma}}\mathcal{L}(\mathbf{Z},\boldsymbol{\gamma}_0)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{\mathcal{I}}_0)$, we can invert system (S18) and apply Slutsky's theorem to get a sandwich variance,

$$\sqrt{M}\begin{bmatrix}\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\\ \hat{\boldsymbol{\lambda}}\end{bmatrix}\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\end{bmatrix}\right),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is the $r\times r$ matrix $\boldsymbol{\mathcal{I}}_0^{-1}(\mathrm{I}_r-\mathbf{H}_0^{\mathrm{T}}(\mathbf{H}_0\boldsymbol{\mathcal{I}}_0^{-1}\mathbf{H}_0^{\mathrm{T}})^{-1}\mathbf{H}_0\boldsymbol{\mathcal{I}}_0^{-1})$ and $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}$ is the $s\times s$ matrix $(\mathbf{H}_0\boldsymbol{\mathcal{I}}_0^{-1}\mathbf{H}_0^{\mathrm{T}})^{-1}$. The rank of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is $r-s$, which represents the number of structural parameters in the model. This completes the proof of asymptotic normality.

References

Aitchison, J. and S. D. Silvey (1958), "Maximum likelihood estimation of parameters subject to restraints." *Annals of Mathematical Statistics*, 29, 813–828. [1, 2, 3, 5]

Gourieroux, C. and A. Monfort (1995), *Statistics and Econometric Models*, Vol. 1. Cambridge University Press, Cambridge, U.K. [1]

Submitted January, 2014. Final version accepted September, 2014.