Web Appendix

Appendix A contains the proofs and derivations of the main theoretical results. Appendix B provides the full details for the Monte Carlo simulations.

Appendix A: Proofs

Proof that weights may be implicitly defined as a function of the parameter vector. Let $F(\omega)$ be the n dimensional vector of equations (6) that defines the n weights

$$F_i(\omega) = \omega_i - \frac{1}{n}\left(\frac{1}{1+\frac{\delta}{1-\delta}(g_i(\theta)-\mathcal{G}_n(\theta))'W_n\mathcal{G}_n(\theta)}\right). \tag{A-1}$$

Then,

$$\begin{aligned}\frac{\partial F_i(\omega)}{\partial \omega_i} &= 1+\frac{1}{n}\frac{\delta}{1-\delta}\left(\frac{1}{1+\frac{\delta}{1-\delta}(g_i(\theta)-\mathcal{G}_n(\theta))'W_n\mathcal{G}_n(\theta)}\right)^2 \\ &\quad \times \left(g_i(\theta)'W_n g_i(\theta) - 2g_i(\theta)'W_n\mathcal{G}_n(\theta)\right) \qquad \text{(A-2)}\\ \frac{\partial F_i(\omega)}{\partial \omega_j} &= \frac{1}{n}\frac{\delta}{1-\delta}\left(\frac{1}{1+\frac{\delta}{1-\delta}(g_i(\theta)-\mathcal{G}_n(\theta))'W_n\mathcal{G}_n(\theta)}\right)^2 \\ &\quad \times \left(g_i(\theta)'W_n g_j(\theta) - 2g_j(\theta)'W_n\mathcal{G}_n(\theta)\right) \qquad \text{(A-3)}\end{aligned}$$

As $n\to\infty$, the matrix $\partial F(\omega)/\partial\omega \xrightarrow{p} \mathbb{I}_n$, which is invertible. By the inverse function theorem, $\omega=\omega(\theta)$ is a continuous differentiable function of θ.

Proof of Theorem 1. First, by the law of large numbers and central limit theorem $V_n(\theta_0)$ and $M_n(\theta_0)$ are consistent estimates of Σ and M respectively such that $V_n(\theta_0)=\Sigma+\mathcal{O}(n^{-\frac{1}{2}})$ and $M_n(\theta_0)=M+\mathcal{O}(n^{-\frac{1}{2}})$.

By equation (21)

$$n\left(\omega_i(\theta_0)-n^{-1}\right) = -\frac{y}{1+y} \tag{A-4}$$

where

$$y=\delta g_i(\theta_0)\mathcal{S}_n(\theta_0)G_n(\theta_0)-\delta(1-\delta)G_n(\theta_0)'\mathcal{S}_n(\theta_0)W_n^{-1}S_n(\theta_0)G_n(\theta_0).$$

By Taylor expansion around $y=0$

$$-\frac{y}{1+y}\approx -y-\frac{1}{2}y^2. \tag{A-5}$$

Suppose $\mathcal{V}_n(\theta_0)=\Sigma+\mathcal{O}(n^{-\frac{1}{2}})$. Then, $\mathcal{S}_n(\theta_0)=\Sigma^{-1}+\mathcal{O}(n^{-\frac{1}{2}})$. Because $G_n(\theta_0)=\mathcal{O}(n^{-\frac{1}{2}})$

$$\delta g_i(\theta_0)\mathcal{S}_n(\theta_0)G_n(\theta_0)=\delta g_i(\theta_0)\left(\Sigma^{-1}+\mathcal{O}\left(n^{-\frac{1}{2}}\right)\right)\mathcal{O}\left(n^{-\frac{1}{2}}\right)=\mathcal{O}\left(n^{-\frac{1}{2}}\right) \tag{A-6}$$

Because the second term of y is quadratic in $G_n(\theta_0)$, it can similarly be shown that

$$\delta(1-\delta)G_n(\theta_0)'\mathcal{S}_n(\theta_0)W_n^{-1}\mathcal{S}_n(\theta_0)G_n(\theta_0) = \mathcal{O}\left(n^{-1}\right). \tag{A-7}$$

Combining the two components of y, we have $n\left(\omega_i(\theta_0) - n^{-1}\right) = \mathcal{O}\left(n^{-\frac{1}{2}}\right)$ if our initial assumption that $\mathcal{V}_n(\theta_0) = \Sigma + \mathcal{O}(n^{-\frac{1}{2}})$ is true. By definition,

$$\begin{aligned}
\mathcal{V}_n(\theta_0) &= \sum_i \omega_i(\theta_0)g_i(\theta_0)g_i(\theta_0)' - \mathcal{G}_n(\theta_0)\mathcal{G}_n(\theta_0)' \\
&= V_n(\theta_0) + \sum_i \left(\omega_i(\theta_0) - n^{-1}\right)v_i(\theta_0) - (1-\delta)^2 G_n(\theta_0)'\mathcal{S}_n(\theta_0)W_n^{-2}\mathcal{S}_n(\theta_0)G_n(\theta_0) \\
&= \left(\Sigma + \mathcal{O}\left(n^{-\frac{1}{2}}\right)\right) + \mathcal{O}\left(n^{-\frac{1}{2}}\right) + \mathcal{O}\left(n^{-1}\right) \\
&= \Sigma + \mathcal{O}\left(n^{-\frac{1}{2}}\right)
\end{aligned} \tag{A-8}$$

Hence, the assumed order of $\mathcal{V}_n(\theta_0)$ is verified to be correct. Thus, part (iii) of the Theorem is proved directly, and the proof for part (i) is now complete. Further, we now have $\mathcal{S}_n(\theta_0) = \Sigma + \mathcal{O}\left(n^{-\frac{1}{2}}\right)$. Part (ii) is proved as follows

$$\begin{aligned}
\sqrt{n}\mathcal{G}_n(\theta_0) &= (1-\delta)W_n^{-1}\mathcal{S}_n(\theta_0)\sqrt{n}G_n(\theta_0) \\
&= (1-\delta)\left(\Sigma + \mathcal{O}\left(n^{-\frac{1}{2}}\right)\right)\left(\Sigma^{-1} + \mathcal{O}\left(n^{-\frac{1}{2}}\right)\right)\sqrt{n}G_n(\theta_0) \\
&= (1-\delta)\Sigma\Sigma^{-1}\sqrt{n}G_n(\theta_0) + \mathcal{O}\left(n^{-\frac{1}{2}}\right).
\end{aligned} \tag{A-9}$$

Simplifying the last equation in the series completes the proof. Finally, part (iv) of the Theorem is proved using part (i)

$$\begin{aligned}
\mathcal{M}_n(\theta_0) &= \sum_i \omega_i(\theta_0)m_i(\theta_0) \\
&= M_n(\theta_0) + \frac{1}{n}\sum_i n\left(\omega_i(\theta_0) - n^{-1}\right)m_i(\theta_0) \\
&= M + \mathcal{O}\left(n^{-\frac{1}{2}}\right) + \frac{1}{n}\sum_i \mathcal{O}\left(n^{-\frac{1}{2}}\right)m_i(\theta_0).
\end{aligned} \tag{A-10}$$

Simplifying the final equation in the series finishes the proof.

Proof of Theorem 2. The form of the proof follows Qin and Lawless (1994). Denote $\theta = \theta_0 + un^{-\varphi}$, for $\theta \in \{\theta | \|\theta - \theta_0\| = n^{-\varphi}\}$, where $\|u\| = 1$. The objective function may be rewritten

$$\begin{aligned} \underset{\text{PMM}}{Q_n}(\theta) =& n(1-\delta)G_n(\theta)'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)G_n(\theta) \\ &+ \frac{2}{\delta}\sum_i \ln\left(1 + \delta g_i(\theta)'\mathcal{S}_n(\theta)G_n(\theta) - \delta(1-\delta)G_n(\theta)'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)G_n(\theta)\right) \end{aligned} \tag{A-11}$$

Taylor approximating the log term by $\ln(1+x) = x - \frac{1}{2}x^2 + o(x^2)$, the objective function is approximated by

$$\begin{aligned} \underset{\text{PMM}}{Q_n}(\theta) =& n(1-\delta)G_n(\theta)'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)G_n(\theta) + 2nG_n(\theta)'\mathcal{S}_n(\theta)G_n(\theta) \\ &- 2n(1-\delta)G_n(\theta)'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)G_n(\theta) - n\delta G_n(\theta)'\mathcal{S}_n(\theta)V_n(\theta)\mathcal{S}_n(\theta)G_n(\theta) \\ &+ o(nG_n(\theta)'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)G_n(\theta)) \end{aligned} \tag{A-12}$$

Using the definition of $\mathcal{S}_n(\theta)$, the following expression emerges

$$\underset{\text{PMM}}{Q_n}(\theta) = nG_n(\theta)'\mathcal{S}_n(\theta)G_n(\theta) + o(nG_n(\theta)'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)G_n(\theta)). \tag{A-13}$$

By Taylor expansion around θ_0, we have (uniformly for u),

$$\begin{aligned} \underset{\text{PMM}}{Q_n}(\theta) =& n\left[G_n(\theta_0) + M_n(\theta_0)un^{-\varphi)}\right]'\mathcal{S}_n(\theta)\left[G_n(\theta_0) + M_n(\theta_0)un^{-\varphi)}\right] \\ &+ o\left(n\left[G_n(\theta_0) + M_n(\theta_0)un^{-\varphi)}\right]'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)\left[G_n(\theta_0) + M_n(\theta_0)un^{-\varphi)}\right]\right). \end{aligned} \tag{A-14}$$

By the law of iterated logarithms and $\varphi < \frac{1}{2}$,

$$\begin{aligned} \underset{\text{PMM}}{Q_n}(\theta) =& n\left[\mathcal{O}(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) + E[M_n(\theta_0)]un^{-\varphi)}\right]' \times \left[\delta E[v_i(\theta_0)] + (1-\delta)W_n^{-1}\right]^{-1} \\ &\times \left[\mathcal{O}(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) + E[M_n(\theta_0)]un^{-\varphi)}\right] + o\left(n^{1-2\varphi}\right) \\ &\geq (c-\varepsilon)n^{1-2\varphi}, \end{aligned} \tag{A-15}$$

where $c - \varepsilon \geq 0$ and c is the smallest eigenvalue of

$$E\left(m_i(\theta_0)\right)'\left[\delta E[v_i(\theta_0)] + (1-\delta)W_n^{-1}\right]^{-1} E\left(m_i(\theta_0)\right). \tag{A-16}$$

Substituting θ_0 into equation (A-13)

$$\begin{aligned} \underset{\text{PMM}}{Q_n}(\theta_0) &= nG_n(\theta_0)'\mathcal{S}_n(\theta_0)G_n(\theta_0) + o(nG_n(\theta_0)'\mathcal{S}_n(\theta_0)W_n^{-1}\mathcal{S}_n(\theta_0)G_n(\theta_0)) \\ &= \mathcal{O}(\log\log n) + o(1) \\ &= \mathcal{O}(\log\log n). \end{aligned} \tag{A-17}$$

Because the objective function is continuous around θ as θ belongs to the ball $\|\theta - \theta_0\| \leq n^{-\varphi}$ and on the surface of the ball the objective function is order $\mathcal{O}(n^{1-2\varphi})$ while the order of the objective function at the population parameter value is $\mathcal{O}(\log\log n)$, the objective function achieves its minimum value within the interior of the ball.

Proof of Theorem 3. We follow Newey and McFadden (1994)'s proof of asymptotic normality for GMM. By assumptions (1) and (2), with probability approaching one the first-order conditions $\mathcal{M}_n(\hat{\theta})'W_n\mathcal{G}_n(\hat{\theta}) = 0$ are satisfied. Expand $G_n(\theta)$ around θ_0 to obtain

$$G_n(\hat{\theta}) = G_n(\theta_0) + M_n(\bar{\theta})'(\hat{\theta} - \theta_0), \tag{A-18}$$

where $\bar{\theta}$ represents a mean value. Substitute in the relationship

$$G_n(\hat{\theta}) = \frac{1}{1-\delta}\mathcal{S}_n(\hat{\theta})^{-1}W_n\mathcal{G}_n(\hat{\theta}) \tag{A-19}$$

and multiply by $\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})$ to obtain

$$0 = \frac{1}{1-\delta}\mathcal{M}_n(\hat{\theta})'W_n\mathcal{G}_n(\hat{\theta}) = \mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})G_n(\theta_0) + \mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})M_n(\bar{\theta})(\hat{\theta} - \theta_0). \tag{A-20}$$

Rearrange to obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left(\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})M_n(\bar{\theta})\right)^{-1}\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})G_n\left(\theta_0\right). \tag{A-21}$$

By assumption (4) and $\hat{\theta} \xrightarrow{p} \theta_0$, with probability approaching one

$$\begin{aligned} &\|M_n(\bar{\theta}) - M\| \leq \|M_n(\bar{\theta}) - M(\bar{\theta})\| + \|M(\bar{\theta}) - M\| \\ &\leq \sup_{\theta\in\Theta}\|M_n(\theta) - M(\theta)\| + \|M(\bar{\theta}) - M\| \xrightarrow{p} 0. \end{aligned} \tag{A-22}$$

Also by assumption (4), $\hat{\theta} \xrightarrow{p} \theta_0$, and Theorem 1 *(iv)* with probability approaching one

$$\begin{aligned} \|\mathcal{M}_n(\hat{\theta}) - M\| &\leq \|\mathcal{M}_n(\hat{\theta}) - \mathcal{M}(\hat{\theta})\| + \|\mathcal{M}(\hat{\theta}) - \mathcal{M}\| + \|\mathcal{M} - M\| \\ &\leq \sup_{\theta\in\Theta}\|\mathcal{M}_n(\theta) - \mathcal{M}(\theta)\| + \|\mathcal{M}(\hat{\theta}) - \mathcal{M}^*\| + \|\mathcal{M} - M\| \xrightarrow{p} 0. \end{aligned} \tag{A-23}$$

Applying assumptions (3), (6), and the Slutsky theorem yields the result.

Proof of Theorem 4. Hansen and Newey and McFadden show that when the asymptotic variance of $\sqrt{n}(\hat{\theta}-\theta_0)$ takes the form $\left((M'\mathcal{S}M)^{-1}M'\mathcal{S}\Sigma\mathcal{S}M(M'\mathcal{S}M)^{-1}\right)$, the minimum variance is obtained when $\mathcal{S} = \Sigma^{-1}$. Because $W_n \xrightarrow{p} \Sigma^{-1}$ by assumption and $\mathcal{V}_n(\theta_0) \xrightarrow{p} \Sigma$, we verify that $\mathcal{S}_n(\theta_0) \xrightarrow{p} \Sigma^{-1}$ the asymptotic variance reduces to $(M'\mathcal{S}M)^{-1}$.

Proof of Theorem 5. Equation (6) may be rewritten

$$n^{-1} = \omega_i + \omega_i \frac{\delta}{1-\delta} \left(g_i(\theta) - \mathcal{G}_n(\theta)\right)' W_n \mathcal{G}_n(\theta). \tag{A-24}$$

Multiply (A-24) by $g_i(\theta)$ and sum over i to obtain

$$\begin{aligned} G_n(\theta) &= \mathcal{G}_n(\theta) + \frac{\delta}{1-\delta}\left(\sum_i \omega_i g_i(\theta) g_i(\theta)' - \mathcal{G}_n(\theta)\mathcal{G}_n(\theta)'\right) W \mathcal{G}_n(\theta) \\ &= \mathcal{G}_n(\theta) + \frac{\delta}{1-\delta}\mathcal{V}_n(\theta) W_n \mathcal{G}_n(\theta). \end{aligned} \tag{A-25}$$

Rearrange to obtain

$$\begin{aligned} \mathcal{G}_n(\theta) &= (1-\delta)\left(\delta \mathcal{V}_n(\theta) W_n + (1-\delta)\mathbb{I}_m\right)^{-1} G_n(\theta) \qquad \text{(A-26)} \\ &= (1-\delta)\left(\delta W_n \mathcal{V}_n(\theta) + (1-\delta)\mathbb{I}_m\right)^{-1} G_n(\theta) \\ &= (1-\delta) W_n^{-1}\left(\delta \mathcal{V}_n(\theta) + (1-\delta) W_n^{-1}\right)^{-1} G_n(\theta) \\ &= (1-\delta) W_n^{-1} \mathcal{S}_n(\theta) G_n(\theta), \end{aligned}$$

where the second line is because W_n and $\mathcal{V}_n(\theta)$ are positive-definite.

Proof of Theorem 6. Equations (12), (19), and (21) provide the result. When δ approaches one, PMM's weights limit to those of EL and $\mathcal{S}_n(\theta) \to \mathcal{V}_n^{-1}(\theta)$. When δ approaches zero, PMM's weights limit to the fixed weights of GMM and $\mathcal{S}_n(\theta) \to W_n$.

Proof of Theorem 7. The final form of the higher order asymptotic expansion is the result of a number of linearizations. First, we give the linear approximations of $\mathcal{S}_n(\hat{\theta})$ and $\mathcal{M}_n(\hat{\theta})$, the derivations of which are included at the end of the proof.

$$\begin{aligned} \mathcal{S}_n(\hat{\theta}) \quad = \quad & \Sigma^{-1} - \Sigma^{-1}\left(V_n\left(\widehat{\theta}\right) - \Sigma\right)\Sigma^{-1} \\ & + \Sigma^{-1}\left(\delta^2 \frac{1}{n}\sum_i g_i\left(\widehat{\theta}\right)'^{-1} G_n\left(\widehat{\theta}\right) v_i\left(\widehat{\theta}\right) + (1-\delta)\sum_{j=1}^{k} \widehat{\Gamma}_j G_n\left(\theta_0\right)'\left(\widetilde{\Upsilon} - \Upsilon\right)' e_j\right)\Sigma^{-1} \\ & + \mathcal{O}\left(n^{-1}\right) \end{aligned} \tag{A-27}$$

and

$$\mathcal{M}_n(\hat{\theta}) = M + \left(M_n(\hat{\theta}) - \delta \frac{1}{n} \sum_i m_i(\hat{\theta}) g_i(\hat{\theta})'^{-1} G_n(\hat{\theta}) - M \right) + \mathcal{O}\left(n^{-1}\right), \tag{A-28}$$

where $\hat{\Gamma}_j = \partial V_n(\hat{\theta})/\partial\theta_j$. In each equation, the first term is the limiting value of the respective equation as $n \to \infty$ evaluated at the true parameter value and is $\mathcal{O}(1)$, the final term is $\mathcal{O}\left(n^{-1}\right)$, and the middle term(s) is the estimation error and is $\mathcal{O}\left(n^{-1/2}\right)$. The sample average Jacobian matrix is rewritten as

$$M_n(\hat{\theta}) = M + \left(M_n(\hat{\theta}) - M \right). \tag{A-29}$$

Again the first term is the limiting value as $n \to \infty$ of the estimated Jacobian evaluated at the true parameter value, the second term is the estimation error, which is $\mathcal{O}\left(n^{-1/2}\right)$, and in this case there is no approximation error. The next step is to expand the sample average moment condition $G_n(\hat{\theta})$ around the true value θ_0

$$G_n(\hat{\theta}) = G_n(\theta_0) + M_n(\hat{\theta}) \left(\hat{\theta} - \theta_0\right) + 0.5 H_n(\hat{\theta}) \left[\left(\hat{\theta} - \theta_0\right) \otimes \left(\hat{\theta} - \theta_0\right) \right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right). \tag{A-30}$$

Equation (7) provides the first-order condition that $\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})G_n(\hat{\theta}) = 0$. Multiply (A-30) by $\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})$ to eliminate the left hand side.

$$\begin{aligned} 0 =& \mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})G_n(\theta_0) + \mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})M_n(\hat{\theta}) \left(\hat{\theta} - \theta_0\right) \\ &+ 0.5\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})H_n(\hat{\theta}) \left[\left(\hat{\theta} - \theta_0\right) \otimes \left(\hat{\theta} - \theta_0\right) \right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right) \end{aligned} \tag{A-31}$$

Define

$$\Omega_n(\hat{\theta}) \equiv \mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})M_n(\hat{\theta}) \qquad \Omega_n^{-1}(\hat{\theta})\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta}) \tag{A-32}$$

and rearrange equation (A-31) to obtain

$$\hat{\theta} - \theta_0 = -\Upsilon_n(\hat{\theta})G_n(\theta_0) - 0.5\Upsilon_n(\hat{\theta})H_n(\hat{\theta}) \left[\left(\hat{\theta} - \theta_0\right) \otimes \left(\hat{\theta} - \theta_0\right) \right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right). \tag{A-33}$$

Next, approximate $\Omega_n^{-1}(\hat{\theta})$ with a two-term Taylor Expansion

$$\begin{aligned}
\Omega_n^{-1}(\hat{\theta}) &= \Omega^{-1} - \Omega^{-1}\left(\Omega_n(\hat{\theta}) - \Omega\right)\Omega^{-1} + \mathcal{O}\left(n^{-1}\right) \\
&= \Omega^{-1} - \Omega^{-1}\left(\mathcal{M}_n(\hat{\theta}) - M\right)^{\prime -1} M + M'\left(\mathcal{S}_n(\hat{\theta}) - \Sigma^{-1}\right) M\Omega^{-1} \\
&\quad - \Omega^{-1}\left(M'^{-1}\left(M_n(\hat{\theta}) - M\right)\right)\Omega^{-1} + \mathcal{O}\left(n^{-1}\right) \\
&= \Omega^{-1} - \Omega^{-1}\left(M_n(\hat{\theta}) - \delta\frac{1}{n}\sum_i m_i(\hat{\theta})g_i(\hat{\theta})'^{-1}G_n(\hat{\theta}) - M\right)^{\prime -1} M\Omega^{-1} \\
&\quad + \Upsilon\left(V_n(\hat{\theta}) - \Sigma\right)\Sigma^{-1}M\Omega^{-1} \\
&\quad - \Upsilon\left(\delta^2\frac{1}{n}\sum_i g_i(\hat{\theta})'^{-1}G_n(\hat{\theta})v_i(\hat{\theta}) + (1-\delta)\sum_{j=1}^{k}\hat{\Gamma}_j G_n(\theta_0)'(\tilde{\Upsilon} - \Upsilon)'e_j\right)\Sigma^{-1}M\Omega^{-1} \\
&\quad - \Upsilon\left(M_n(\hat{\theta}) - M\right)\Omega^{-1} + \mathcal{O}\left(n^{-1}\right).
\end{aligned} \tag{A-34}$$

The final equation in the series is the result of substituting in the approximations provided by equations (A-27) to (A-29). The first term is $\mathcal{O}(1)$ and the next three terms are $\mathcal{O}\left(n^{-1/2}\right)$.

In equation (A-33), the first term includes $G_n(\theta_0)$ which is $\mathcal{O}\left(n^{-1/2}\right)$ and the second term includes $(\hat{\theta} - \theta_0) \otimes (\hat{\theta} - \theta_0)$ which is $\mathcal{O}\left(n^{-1}\right)$. Hence, because we are investigating the $\mathcal{O}\left(n^{-1}\right)$ properties of the estimator $\hat{\theta}$, for the first term, we only require the $\mathcal{O}\left(n^{-1/2}\right)$ components of $\Omega_n^{-1}(\hat{\theta})$, $\mathcal{M}_n(\hat{\theta})$, and $\mathcal{S}_n(\hat{\theta})$. For the second term, only their limiting values as $n \to \infty$ evaluated at the true parameter value θ_0 are needed. The rest of the terms are

$\mathcal{O}\left(n^{-3/2}\right)$. After substituting in the relevant terms

$$
\begin{aligned}
\hat{\theta} - \theta_0 = & - \Upsilon G_n(\theta_0) \\
& + \Omega^{-1}\left(M_n(\hat{\theta}) - \delta\frac{1}{n}\sum_i m_i(\hat{\theta})g_i(\hat{\theta})'^{-1}G_n(\hat{\theta}) - M\right)'^{-1} M\Omega^{-1}M'^{-1}G_n(\theta_0) \\
& - \Upsilon\left(V_n(\hat{\theta}) - \delta^2\frac{1}{n}\sum_i g_i(\hat{\theta})'^{-1}G_n(\hat{\theta})v_i(\hat{\theta}) - \Sigma\right)\Sigma^{-1}M\Omega^{-1}M'^{-1}G_n(\theta_0) \\
& + \Upsilon\left((1-\delta)\sum_{j=1}^{k}\hat{\Gamma}_j G_n(\theta_0)'(\tilde{\Upsilon} - \Upsilon)'e_j\right)\Sigma^{-1}M\Omega^{-1}M'^{-1}G_n(\theta_0) \\
& + \Upsilon\left(M_n(\hat{\theta}) - M\right)\Upsilon G_n(\theta_0) \\
& - \Omega^{-1}\left(M_n(\hat{\theta}) - \delta\frac{1}{n}\sum_i m_i(\hat{\theta})g_i(\hat{\theta})'^{-1}G_n(\hat{\theta}) - M\right)\Sigma^{-1}G_n(\theta_0) \\
& + \Upsilon\left(V_n(\hat{\theta}) - \delta^2\frac{1}{n}\sum_i g_i(\hat{\theta})'^{-1}G_n(\hat{\theta})v_i(\hat{\theta}) - \Sigma\right)\Sigma^{-1}G_n(\theta_0) \\
& - \Upsilon\left((1-\delta)\sum_{j=1}^{k}\hat{\Gamma}_j G_n(\theta_0)'(\tilde{\Upsilon} - \Upsilon)'e_j\right)\Sigma^{-1}G_n(\theta_0) \\
& - 0.5\Upsilon H_n(\hat{\theta})\left[\left(\hat{\theta} - \theta_0\right)\otimes\left(\hat{\theta} - \theta_0\right)\right] \\
& + \mathcal{O}\left(n^{-\frac{3}{2}}\right).
\end{aligned} \tag{A-35}
$$

The first line of (A-35) is $\mathcal{O}\left(n^{-1/2}\right)$ and lines two through nine are $\mathcal{O}\left(n^{-1}\right)$. Lines two through five are from the $\mathcal{O}\left(n^{-1/2}\right)$ component of the expansion of $\Omega_n^{-1}(\hat{\theta})$. The second line is due to $\mathcal{M}_n(\hat{\theta})'$. Lines three and four are attributed to $\mathcal{S}_n(\hat{\theta})$. The fifth line is from $M_n(\hat{\theta})$. In addition, $\mathcal{O}\left(n^{-1/2}\right)$ terms are embedded in the $\mathcal{M}_n(\hat{\theta})'$ and $\mathcal{S}_n(\hat{\theta})$, the other two components of $\Upsilon_n(\hat{\theta})$. Line six is due to $\mathcal{M}_n(\hat{\theta})'$ and lines seven and eight are from $\mathcal{S}_n(\hat{\theta})$. Finally, the ninth line is the third term in the expansion of $G_n(\hat{\theta})$. In order to calculate the $\mathcal{O}\left(n^{-1}\right)$ bias, take expectations of (A-35) and rearrange to obtain

$$
\begin{aligned}
\mathbb{E}\left[\hat{\theta} - \theta_0\right] = \quad & \Upsilon\mathbb{E}\left[M_n(\hat{\theta})\Upsilon G_n(\theta_0)\right] - 0.5\Upsilon H\mathbb{E}\left[\left(\hat{\theta} - \theta_0\right)\otimes\left(\hat{\theta} - \theta_0\right)\right] \\
& - \Omega^{-1}\mathbb{E}\left[\left(M_n(\hat{\theta}) - \delta\frac{1}{n}\sum_i m_i(\hat{\theta})g_i(\hat{\theta})'^{-1}G_n(\hat{\theta})\right)'^{-\frac{1}{2}}\mathbf{P}_{\overline{M}}^{\perp}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right] \\
& + \Upsilon\mathbb{E}\left[\left(V_n(\hat{\theta}) - \delta^2\frac{1}{n}\sum_i g_i(\hat{\theta})'^{-1}G_n(\hat{\theta})v_i(\hat{\theta})\right)\Sigma^{-\frac{1}{2}}\mathbf{P}_{\overline{M}}^{\perp}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right] \\
& - \Upsilon\mathbb{E}\left[\left((1-\delta)\sum_{j=1}^{k}\hat{\Gamma}_j G_n(\theta_0)'(\tilde{\Upsilon} - \Upsilon)'e_j\right)\Sigma^{-\frac{1}{2}}\mathbf{P}_{\overline{M}}^{\perp}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right] \\
& + \mathcal{O}\left(n^{-\frac{3}{2}}\right),
\end{aligned} \tag{A-36}
$$

where $\mathbf{P}^{\perp}_{\overline{M}} \equiv \mathbb{I}_m - \Sigma^{-\frac{1}{2}} M(M'\Sigma^{-1}M)^{-1}\Sigma^{-\frac{1}{2}} M$ is a projection matrix orthogonal to the space spanned by the asymptotic normalized Jacobian. Before proceeding, note

$$g_i(\hat{\theta}) = g_i(\theta_0) + \mathcal{O}\left(n^{-\frac{1}{2}}\right) \tag{A-37}$$

$$m_i(\hat{\theta}) = m_i(\theta_0) + \mathcal{O}\left(n^{-\frac{1}{2}}\right) \tag{A-38}$$

$$G_n(\hat{\theta}) = G_n(\theta_0) + M'\left(\hat{\theta} - \theta_0\right) + \left(M_n(\hat{\theta}) - M\right)'\left(\hat{\theta} - \theta_0\right) + \mathcal{O}\left(n^{-1}\right) \tag{A-39}$$

$$nG_n(\theta_0)G_n(\theta_0)'\Sigma^{-1} = \mathbb{I}_m + \underbrace{\left(nG_n(\theta_0)G_n(\theta_0)'\Sigma^{-1} - \mathbb{I}_m\right)}_{\mathcal{O}\left(n^{-1/2}\right)}. \tag{A-40}$$

Using the result of the expansion of $G_n(\hat{\theta})$: $\hat{\theta} - \theta_0 = -(M'^{-1}M)^{-1}M'^{-1}G_n(\theta_0) + \mathcal{O}\left(n^{-1}\right)$, the assumption that the observations are independent, and equation (A-38), the first line of (A-36) may be rewritten

$$B_I = n^{-1}\Upsilon\left(\mathbb{E}\left[m_i(\theta_0)\Upsilon g_i(\theta_0)\right] - a\right), \tag{A-41}$$

where a is an $m \times 1$ matrix such that

$$a_j \equiv 0.5\mathrm{tr}\left(\Omega^{-1}\mathbb{E}\left[\frac{\partial^2 g_{ij}(\theta_0)}{\partial\theta\partial\theta'}\right]\right) \qquad (j = 1, \ldots, m), \tag{A-42}$$

and $g_{ij}(\theta_0)$ represents the jth element of $g_i(\theta)$.

The next step is to rewrite the second line of (A-36). The three approximations given by equations (A-37) through (A-39) are substituted into the second line of (A-36). Note that the second term of (A-38) is orthogonal to $\mathbf{P}^{\perp}_{\overline{M}}$ and the third term is $\mathcal{O}\left(n^{-1}\right)$.

$$\begin{aligned}
&-\Omega^{-1}\mathbb{E}\left[\left(M_n(\hat{\theta}) - \delta\frac{1}{n}\sum_i m_i(\hat{\theta})g_i(\hat{\theta})'^{-1}G_n(\hat{\theta})\right)'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right] \\
&\quad = -\Omega^{-1}\mathbb{E}\left[\frac{1}{n}\sum_i m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right] \\
&\qquad + \Omega^{-1}\mathbb{E}\left[\delta\frac{1}{n}\sum_i m_i(\hat{\theta})'g_i(\theta_0)'^{-1}G_n(\theta_0)'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right)
\end{aligned}$$

$$= -\Omega^{-1}\mathbb{E}\left[\frac{1}{n}\sum_i m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right]$$

$$+ \Omega^{-1}\mathbb{E}\left[\delta\frac{1}{n}\sum_i m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)G_n(\theta_0)'^{-1}g_i(\theta_0)\right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right)$$

$$= -\Omega^{-1}\mathbb{E}\left[\frac{1}{n}\sum_i m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right]$$

$$+ \Omega^{-1}\mathbb{E}\left[\delta\frac{1}{n^2}\sum_i m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}\left(nG_n(\theta_0)G_n(\theta_0)'\right)\Sigma^{-1}g_i(\theta_0)\right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right)$$

$$= -\Omega^{-1}\mathbb{E}\left[\frac{1}{n^2}\sum_i m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\left(\mathbb{I}_m - \delta\Sigma^{-\frac{1}{2}}\left(nG_n(\theta_0)G_n(\theta_0)'\right)\Sigma^{-\frac{1}{2}}\right)\Sigma^{-\frac{1}{2}}g_i(\theta_0)\right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right)$$

$$= -\Omega^{-1}\mathbb{E}\left[\frac{1}{n^2}\sum_i m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\left(\mathbb{I}_m - \delta\mathbb{I}_m\right)\Sigma^{-\frac{1}{2}}g_i(\theta_0)\right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right)$$

$$= -n^{-1}\left(1-\delta\right)\Omega^{-1}\mathbb{E}\left[m_i(\hat{\theta})'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}g_i(\theta_0)\right] + \mathcal{O}\left(n^{-\frac{3}{2}}\right)$$

$$= \underbrace{-n^{-1}\left(1-\delta\right)\Omega^{-1}\mathbb{E}\left[m_i(\theta_0)'^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}g_i(\theta_0)\right]}_{(1-\delta)B_M} + \mathcal{O}\left(n^{-\frac{3}{2}}\right)$$

The exact same procedure for the third line of equation (A-36) provides the bias term associated with the estimation of the second moment matrix

$$B_\Sigma = n^{-1}\left(1-\delta^2\right)\Upsilon\mathbb{E}\left[v_i(\theta_0)\Sigma^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}g_i(\theta_0)\right]\ . \tag{A-43}$$

Finally, rewrite the fourth line of equation (A-36)

$$B_W = -\Upsilon\mathbb{E}\left[\left((1-\delta)\sum_{j=1}^{k}\hat{\Gamma}_j G_n(\theta_0)'(\tilde{\Upsilon}-\Upsilon)'e_j\right)\Sigma^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)\right]$$

$$= -\Upsilon\mathbb{E}\left[(1-\delta)\sum_{j=1}^{k}\hat{\Gamma}_j\Sigma^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}G_n(\theta_0)G_n(\theta_0)'(\tilde{\Upsilon}-\Upsilon)'e_j\right]$$

$$= -n^{-1}\Upsilon\left(1-\delta\right)\mathbb{E}\left[\sum_{j=1}^{k}\hat{\Gamma}_j\Sigma^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}\left(nG_n(\theta_0)G_n(\theta_0)'\right)(\tilde{\Upsilon}-\Upsilon)'e_j\right] \tag{A-44}$$

The second line moves the scalar value $G_n(\theta_0)'(\tilde{\Upsilon}-\Upsilon)'e_j$. The third line multiplies and divides by n and takes $(1-\delta)$ outside the expectation. Then, by noting $\mathbf{P}^{\perp}_{\overline{M}}$ is symmetric and by applying the note from the proof of Theorem 4.1 in Newey and Smith (2004) for $\Sigma^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\overline{M}}\Sigma^{-\frac{1}{2}}$ and using the approximation given in equation (A-40) for $nG_n(\theta_0)G_n(\theta_0)'$ the result is obtained.

$$
\begin{aligned}
&= -n^{-1}\Upsilon\,(1-\delta)\,\mathbb{E}\left[\sum_{j=1}^{k}\hat{\Gamma}_j\Sigma^{-\frac{1}{2}}\mathbf{P}_{\overline{M}}^{\perp}\Sigma^{-\frac{1}{2}}\left(nG_n(\theta_0)G_n(\theta_0)'\right)(\tilde{\Upsilon}-\Upsilon)'e_j\right]\\
&= -n^{-1}\,(1-\delta)\,\Upsilon\mathbb{E}\left[\sum_{j=1}^{k}\hat{\Gamma}_j(\tilde{\Upsilon}-\Upsilon)'e_j\right]+\mathcal{O}\left(n^{-\frac{3}{2}}\right) \qquad \text{(A-45)}
\end{aligned}
$$

The final results

$$
\begin{aligned}
\text{Bias}\left(\hat{\theta}_{\text{PMM}}\right) &= B_I + B_M + B_\Sigma + B_W \qquad \text{(A-46)}\\
B_I &= n^{-1}\Upsilon\left(\mathbb{E}\left[m_i(\theta_0)\Upsilon g_i(\theta_0)\right]-a\right)\\
B_M &= -n^{-1}\,(1-\delta)\,\Omega^{-1}\mathbb{E}\left[m_i(\theta_0)'^{-\frac{1}{2}}\mathbf{P}_{\overline{M}}^{\perp}\Sigma^{-\frac{1}{2}}g_i(\theta_0)\right]\\
B_\Sigma &= n^{-1}\left(1-\delta^2\right)\Upsilon\mathbb{E}\left[v_i(\theta_0)\Sigma^{-\frac{1}{2}}\mathbf{P}_{\overline{M}}^{\perp}\Sigma^{-\frac{1}{2}}g_i(\theta_0)\right]\\
B_W &= -n^{-1}\,(1-\delta)\,\Upsilon\mathbb{E}\left[\sum_{j=1}^{k}\hat{\Gamma}_j\right](\tilde{\Upsilon}-\Upsilon)'e_j.
\end{aligned}
$$

The Three Linearizations: $n\omega_i(\hat{\theta})$, $\mathcal{S}_n(\hat{\theta})$, and $\mathcal{M}_n(\hat{\theta})$

The first step in the two linearizations is to linearize the probability weights $\omega_i(\hat{\theta})$. Recall the definition of the optimal PMM weights conditional on θ is

$$
\omega_i(\theta) = \frac{1}{n}\frac{1}{1+\delta\underbrace{g_i(\theta)'\mathcal{S}_n(\theta)G_n(\theta)}_{\mathcal{O}\left(n^{-1/2}\right)}-\frac{\kappa}{(1+\kappa)^2}\underbrace{G_n(\theta)'\mathcal{S}_n(\theta)W_n^{-1}\mathcal{S}_n(\theta)G_n(\theta)}_{\mathcal{O}(n^{-1})}}. \qquad \text{(A-47)}
$$

A geometric expansion provides the linearized weights

$$
\omega_i(\theta) = \frac{1}{n} - \frac{1}{n}\delta g_i(\theta)'\mathcal{S}_n(\theta)G_n(\theta) + \mathcal{O}\left(n^{-2}\right). \qquad \text{(A-48)}
$$

Equation (A-48) can be further refined because $\mathcal{S}_n(\theta)$ is an estimate of the second moment matrix. Recall the definition of $\mathcal{S}_n(\theta)$

$$
\mathcal{S}_n(\hat{\theta}) = \left(\delta\mathcal{V}_n(\hat{\theta}) + (1-\delta)\,W_n^{-1}\right)^{-1}. \qquad \text{(A-49)}
$$

Because W_n^{-1} is a consistent estimate of Σ, $\mathcal{S}_n(\hat{\theta}$ may be approximated by

$$
\mathcal{S}_n(\hat{\theta}) = \Sigma^{-1} - \Sigma^{-1}\underbrace{\left(\delta\mathcal{V}_n(\hat{\theta}) + (1-\delta)\,W_n^{-1} - \Sigma\right)}_{\mathcal{O}\left(n^{-1/2}\right)}\Sigma^{-1}. \qquad \text{(A-50)}
$$

Before continuing with the approximation $\mathcal{S}_n(\hat{\theta})$, we substitute equation (A-50) into (4) to obtain

$$\omega_i(\hat{\theta}) = \frac{1}{n} - \frac{1}{n}\delta g_i(\hat{\theta})'^{-1} G_n(\hat{\theta}) + \mathcal{O}\left(n^{-2}\right). \tag{A-51}$$

Next, use equation (A-51) to approximate $\mathcal{V}_n(\hat{\theta})$

$$\begin{aligned}\mathcal{V}_n(\hat{\theta}) &= \sum_i \omega_i(\hat{\theta}) v_i(\hat{\theta}) \\ &= V_n(\hat{\theta}) - \delta \frac{1}{n} \sum_i v_i(\hat{\theta}) g_i(\hat{\theta})'^{-1} G_n(\hat{\theta}) + \mathcal{O}\left(n^{-1}\right) \end{aligned} \tag{A-52}$$

and rewrite $W_n^{-1} = V_n(\tilde{\theta})$, where $\tilde{\theta}$ is a first-round estimator of θ, as

$$\begin{aligned} W_n^{-1} &= V_n(\hat{\theta}) + \left(V_n(\tilde{\theta}) - V_n(\hat{\theta})\right) \\ V_n(\tilde{\theta}) &= V_n(\theta_0) + \sum_{j=1}^{k} \frac{\partial V_n(\theta_0)}{\partial \theta_j} \left(\tilde{\theta}_j - \theta_{0j}\right) + \mathcal{O}\left(n^{-1}\right) \\ &= V_n(\theta_0) - \sum_{j=1}^{k} \frac{\partial V_n(\theta_0)}{\partial \theta_j} e_j \left(M'\tilde{\mathcal{S}}M\right)^{-1} M'\tilde{\mathcal{S}} G_n(\theta_0) + \mathcal{O}\left(n^{-1}\right) \end{aligned} \tag{A-53}$$

$$\begin{aligned} V_n(\hat{\theta}) &= V_n(\theta_0) + \sum_{j=1}^{k} \frac{\partial V_n(\theta_0)}{\partial \theta_j} \left(\hat{\theta}_j - \theta_{0j}\right) + \mathcal{O}\left(n^{-1}\right) \\ &= V_n(\theta_0) - \sum_{j=1}^{k} \frac{\partial V_n(\theta_0)}{\partial \theta_j} e_j \left(M'\hat{\mathcal{S}}M\right)^{-1} M'\hat{\mathcal{S}} G_n(\theta_0) + \mathcal{O}\left(n^{-1}\right) \\ W_n^{-1} &= V_n(\hat{\theta}) - \sum_{j=1}^{k} \frac{\partial V_n(\theta_0)}{\partial \theta_j} e_j' \left(\tilde{\Upsilon} - \hat{\Upsilon}\right) G_n(\theta_0) + \mathcal{O}\left(n^{-1}\right) \\ &= V_n(\hat{\theta}) - \sum_{j=1}^{k} \frac{\partial V_n(\theta_0)}{\partial \theta_j} G_n(\theta_0)' \left(\tilde{\Upsilon} - \Upsilon\right)' e_j + \mathcal{O}\left(n^{-1}\right). \end{aligned} \tag{A-54}$$

The last line is because $\hat{\Upsilon} = \Upsilon + \mathcal{O}\left(n^{-1/2}\right)$. This is not the case for $\tilde{\Upsilon}$, because $\tilde{\mathcal{S}}$ is a function of the the first round weighting matrix $\tilde{W}$, which is not necessarily a consistent estimate of Σ^{-1}. Substituting equations (A-52) and (A-54) into (A-50) provides the result

$$\begin{aligned} \mathcal{S}_n(\hat{\theta}) =& \Sigma^{-1} - \Sigma^{-1}\left(V_n(\hat{\theta}) - \Sigma\right)\Sigma^{-1} \\ &+ \Sigma^{-1}\left((\delta)^2 \frac{1}{n} \sum_i g_i(\hat{\theta})'^{-1} G_n(\hat{\theta}) v_i(\hat{\theta}) + (1-\delta)\sum_{j=1}^{k} \hat{\Gamma}_j G_n(\theta_0)'(\tilde{\Upsilon} - \Upsilon)' e_j \right)\Sigma^{-1} \\ &+ \mathcal{O}\left(n^{-1}\right). \end{aligned} \tag{A-55}$$

Proof of Theorem 8. PMM's objective function has two components, the quadratic penalty and the KLIC penalty.

By Theorem 5, the quadratic component of the objective function is rewritten

$$n(1-\delta)G_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})W_n^{-1}\mathcal{S}_n(\hat{\theta})G_n(\hat{\theta}). \tag{A-56}$$

Taylor approximating logarithms by $ln(1+x) = x - \frac{1}{2}x^2 + \mathcal{O}(x^3)$ and substituting in the optimal PMM weights conditioned on θ given by equation (21) gives

$$\begin{aligned}\ln(n\omega_i(\hat{\theta})) =&\delta g_i(\theta)'\mathcal{S}_n(\hat{\theta})G_n(\theta) - \delta(1-\delta)G_n(\hat{\theta})\mathcal{S}_n(\hat{\theta})W_n^{-1}\mathcal{S}_n(\hat{\theta})G_n(\theta)\\ &- \frac{1}{2}\delta^2 G_n(\hat{\theta})\mathcal{S}_n(\hat{\theta})g_i(\hat{\theta})g_i(\hat{\theta})'\mathcal{S}_n(\hat{\theta})G_n(\hat{\theta}) + \mathcal{O}\left(n^{-\frac{3}{2}}\right).\end{aligned} \tag{A-57}$$

Summing equation (A-57) over i provides

$$\begin{aligned}\sum_i \ln(n\omega_i(\hat{\theta})) =&\delta n G_n(\theta)'\mathcal{S}_n(\hat{\theta})G_n(\theta) - n\delta(1-\delta)G_n(\hat{\theta})\mathcal{S}_n(\hat{\theta})W_n^{-1}\mathcal{S}_n(\hat{\theta})G_n(\theta)\\ &- \frac{n}{2}\delta^2 G_n(\hat{\theta})\mathcal{S}_n(\hat{\theta})V_n(\hat{\theta})\mathcal{S}_n(\hat{\theta})G_n(\hat{\theta}) + \mathcal{O}(n^{-\frac{1}{2}}).\end{aligned} \tag{A-58}$$

Dividing equation (A-56) by $1-\delta$ and equation (A-58) by $-\frac{\delta}{2}$, adding the two components, and simplifying the expression

$$\underset{\text{PMM}}{Q}(\hat{\theta}) = nG_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})G_n(\hat{\theta}) + \mathcal{O}\left(n^{-\frac{1}{2}}\right). \tag{A-59}$$

The quadratic statistic is singular

$$\begin{aligned}nG_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})G_n(\hat{\theta}) &= nG_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}\mathbb{I}_m\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}G_n(\hat{\theta})\\ &= nG_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}\left(\mathbf{P}_{\bar{\mathcal{M}}(\hat{\theta})} + \mathbf{P}^{\perp}_{\bar{\mathcal{M}}(\hat{\theta})}\right)\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}G_n(\hat{\theta})\\ &= nG_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}\mathbf{P}^{\perp}_{\bar{\mathcal{M}}(\hat{\theta})}\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}G_n(\hat{\theta})\\ &= \text{rank}(m-k)\end{aligned} \tag{A-60}$$

where $\mathbf{P}_{\bar{\mathcal{M}}(\hat{\theta})} \equiv \mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}\mathcal{M}_n(\hat{\theta})\left(\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})\mathcal{M}_n(\hat{\theta})\right)^{-1}\mathcal{M}_n(\hat{\theta})'\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}$ is the projection matrix that sets k dimensions of $G_n(\hat{\theta})$ to zero and $\mathcal{S}_n(\hat{\theta})^{\frac{1}{2}}$ is the Cholesky decomposition of $\mathcal{S}_n(\hat{\theta})$. If $W_n \xrightarrow{p} \Sigma^{-1}$ then $\mathcal{S}_n(\hat{\theta}) \xrightarrow{p} \Sigma^{-1}$.

It is asymptotically equivalent to use the properly scaled components of PMM's objective function as test statistics: $n(1-\delta)^{-2}\mathcal{G}_n(\hat{\theta})'W_n^{-1}\mathcal{G}_n(\hat{\theta}) \xrightarrow{d} \chi^2_{m-k}$ and $-2\delta^{-2}\sum_i \ln(n\hat{\omega}_i) \xrightarrow{d} \chi^2_{m-k}$.

Appendix B: Monte Carlo Simulations

The PMM estimation problem can be redefined to achieve a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$, reducing the computational complexity to that of EL. To do so, rewrite equation (6) replacing $\mathcal{G}_n(\theta)$ with the m-dimensional vector λ

$$\omega_i(\lambda, \theta) = \frac{1}{N}\left(\frac{1}{1 + \frac{\delta}{1-\delta}(g_i(\theta) - \lambda)' W_n \lambda}\right). \tag{A-61}$$

Find the λ such that $\lambda = \mathcal{G}_n(\theta)$. First, define

$$f(\lambda, \theta) = \lambda - \sum_i \omega_i(\lambda, \theta) g_i(\theta). \tag{A-62}$$

Then λ may be obtained via the Newton-Raphson iterative algorithm

$$\lambda_{j+1} = \lambda_j - \left[\mathbb{I}_m + n\frac{\delta}{1-\delta}\sum_i \omega_i^2(\lambda_j, \theta) g_i(\theta)\left(g_i(\theta) - 2\lambda_j\right)' W_n\right]^{-1}\left[\lambda_j - \sum_i \omega_i(\lambda_j, \theta) g_i(\theta)\right]. \tag{A-63}$$

Theorem 1 suggests an appropriate starting value for the iteration: $\lambda_0 = (1-\delta)G_n(\theta)$. In practice, we have noticed that setting $\lambda_0 = 0.25(1-\delta)G_n(\theta)$ leads to fewer instances of divergence. The iterative procedure described above constitutes an *inner loop* which must be solved for any given θ yielding $\lambda(\theta)$. The *outer loop* selects θ to minimize the objective function defined by equation (1)

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \quad \frac{1}{1-\delta} n\lambda(\theta)' W \lambda(\theta) + \frac{2}{\delta}\sum_i \ln\left(1 + \frac{\delta}{1-\delta}(g_i(\theta) - \lambda(\theta))' W \lambda(\theta)\right). \tag{A-64}$$

The simulations discussed in Section 3.1 are based on the equations outlined above, the Hall and Horowitz (1996) model detailed in the main text, and sample sizes of $n = 25$, 50, and 100. For the $n = 25$ trials, we consider a 2 moment and 5 moment model. For $n = 50$, we use 2, 5, and 10 moments. For $n = 100$, we use 2, 5, 10, 15, and 20 moments. Each batch of simulations includes 2000 trials. The results are presented in both tabular and graphical format. We calculate the realized estimator mean, volatility, and Root Mean Squared Error (RMSE) for PMM and EL. The size of the rejection region for the relevant test of the overidentifying restrictions with α chosen to be 0.10, 0.05, and 0.01 percent, and the probability of rejecting the population parameter value for α equal to 0.10, 0.05, and 0.01 percent are also calculated. Table 1 reports the results for the 25 and 50 sample estimations and Table 2 reports the results for the 100 sample estimations. For each Monte Carlo experiment, we plot the densities of the realized estimates obtained under EL and PMM estimation. Figures 1 and 2 provide the plots for the 25 observation simulations. Figures 3 through 5 plot the densities for the 50 observation experiments. The densities for estimates obtained under sample size of 100 are plotted in Figures 6 through 10.

Table 1: **Monte Carlo Simulations.** This table presents the results of the Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample sizes of 25 and 50 observations and 2000 trials. The population parameter value θ_0 is 3. $E(\hat{\theta})$ is the sample average of the estimates, $\sigma(\hat{\theta})$ is the sample volatility of the estimates, and $rmse(\hat{\theta})$ is the Root Mean Squared Error of the estimates. χ^2_α and t_α respectively represent the size of the rejection region for the test of overidentifying restrictions and test that the estimate equals the population parameter value at the α level.

Model	$E(\hat{\theta})$	$\sigma(\hat{\theta})$	$rmse(\hat{\theta})$	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.01}$	$t_{0.10}$	$t_{0.05}$	$t_{0.01}$
25 Observations and 2 Moment Conditions									
EL	3.092	0.778	0.783	0.199	0.121	0.036	0.444	0.396	0.317
	(0.017)	(0.020)	(0.020)	(0.009)	(0.007)	(0.004)	(0.011)	(0.011)	(0.010)
PMM	3.110	0.791	0.798	0.223	0.154	0.054	0.443	0.392	0.313
	(0.018)	(0.021)	(0.021)	(0.009)	(0.008)	(0.005)	(0.011)	(0.011)	(0.010)
25 Observations and 5 Moment Conditions									
EL	3.484	0.939	1.056	0.611	0.531	0.390	0.477	0.438	0.378
	(0.021)	(0.026)	(0.028)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
PMM	2.922	0.850	0.854	0.479	0.352	0.153	0.410	0.360	0.280
	(0.019)	(0.020)	(0.021)	(0.011)	(0.011)	(0.008)	(0.011)	(0.011)	(0.010)
50 Observations and 2 Moment Conditions									
EL	3.098	0.417	0.428	0.202	0.136	0.064	0.339	0.276	0.205
	(0.009)	(0.007)	(0.008)	(0.009)	(0.008)	(0.005)	(0.011)	(0.010)	(0.009)
PMM	3.049	0.493	0.495	0.186	0.119	0.037	0.364	0.307	0.232
	(0.011)	(0.012)	(0.012)	(0.009)	(0.007)	(0.004)	(0.011)	(0.010)	(0.009)
50 Observations and 5 Moment Conditions									
EL	3.082	0.461	0.468	0.469	0.359	0.177	0.335	0.280	0.201
	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.009)	(0.011)	(0.010)	(0.009)
PMM	3.154	0.467	0.491	0.472	0.373	0.195	0.352	0.294	0.226
	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.009)	(0.011)	(0.010)	(0.009)
50 Observations and 10 Moment Conditions									
EL	3.504	0.724	0.881	0.793	0.733	0.576	0.468	0.434	0.373
	(0.016)	(0.018)	(0.021)	(0.009)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
PMM	2.882	0.595	0.607	0.727	0.628	0.400	0.310	0.255	0.186
	(0.013)	(0.016)	(0.016)	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)	(0.009)

Table 2: **Monte Carlo Simulations.** This table presents the results of the Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 100 observations and 2000 trials. The population parameter value θ_0 is 3. $E(\hat{\theta})$ is the sample average of the estimates, $\sigma(\hat{\theta})$ is the sample volatility of the estimates, and $rmse(\hat{\theta})$ is the Root Mean Squared Error of the estimates. χ^2_α and t_α respectively represent the size of the rejection region for the test of overidentifying restrictions and test that the estimate equals the population parameter value at the α level.

Model	$E(\hat{\theta})$	$\sigma(\hat{\theta})$	$rmse(\hat{\theta})$	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.01}$	$t_{0.10}$	$t_{0.05}$	$t_{0.01}$
100 Observations and 2 Moment Conditions									
EL	3.058	0.299	0.304	0.186	0.119	0.039	0.271	0.222	0.153
	(0.007)	(0.005)	(0.006)	(0.009)	(0.007)	(0.004)	(0.010)	(0.009)	(0.008)
PMM	3.050	0.319	0.323	0.182	0.114	0.038	0.284	0.232	0.161
	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)	(0.004)	(0.010)	(0.009)	(0.008)
100 Observations and 5 Moment Conditions									
EL	3.148	0.322	0.354	0.404	0.311	0.169	0.297	0.246	0.165
	(0.007)	(0.007)	(0.008)	(0.011)	(0.010)	(0.008)	(0.010)	(0.010)	(0.008)
PMM	3.093	0.318	0.331	0.433	0.325	0.170	0.265	0.210	0.139
	(0.007)	(0.007)	(0.007)	(0.011)	(0.010)	(0.008)	(0.010)	(0.009)	(0.008)
100 Observations and 10 Moment Conditions									
EL	3.054	0.320	0.324	0.700	0.605	0.413	0.214	0.162	0.107
	(0.007)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.009)	(0.008)	(0.007)
PMM	3.144	0.327	0.357	0.679	0.592	0.402	0.267	0.210	0.144
	(0.007)	(0.006)	(0.007)	(0.010)	(0.011)	(0.011)	(0.010)	(0.009)	(0.008)
100 Observations and 15 Moment Conditions									
EL	3.376	0.412	0.558	0.846	0.783	0.642	0.440	0.381	0.296
	(0.009)	(0.009)	(0.010)	(0.008)	(0.009)	(0.011)	(0.011)	(0.011)	(0.010)
PMM	3.001	0.340	0.340	0.848	0.776	0.581	0.189	0.143	0.084
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.011)	(0.009)	(0.008)	(0.006)
100 Observations and 20 Moment Conditions									
EL	3.442	0.454	0.634	0.935	0.895	0.805	0.469	0.403	0.313
	(0.010)	(0.011)	(0.012)	(0.006)	(0.007)	(0.009)	(0.011)	(0.011)	(0.010)
PMM	2.885	0.370	0.387	0.919	0.871	0.712	0.177	0.125	0.076
	(0.008)	(0.008)	(0.009)	(0.006)	(0.007)	(0.010)	(0.009)	(0.007)	(0.006)

Figure 1: **Kernel Density Plot for 25 Observations and 2 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 25 observations, 2 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

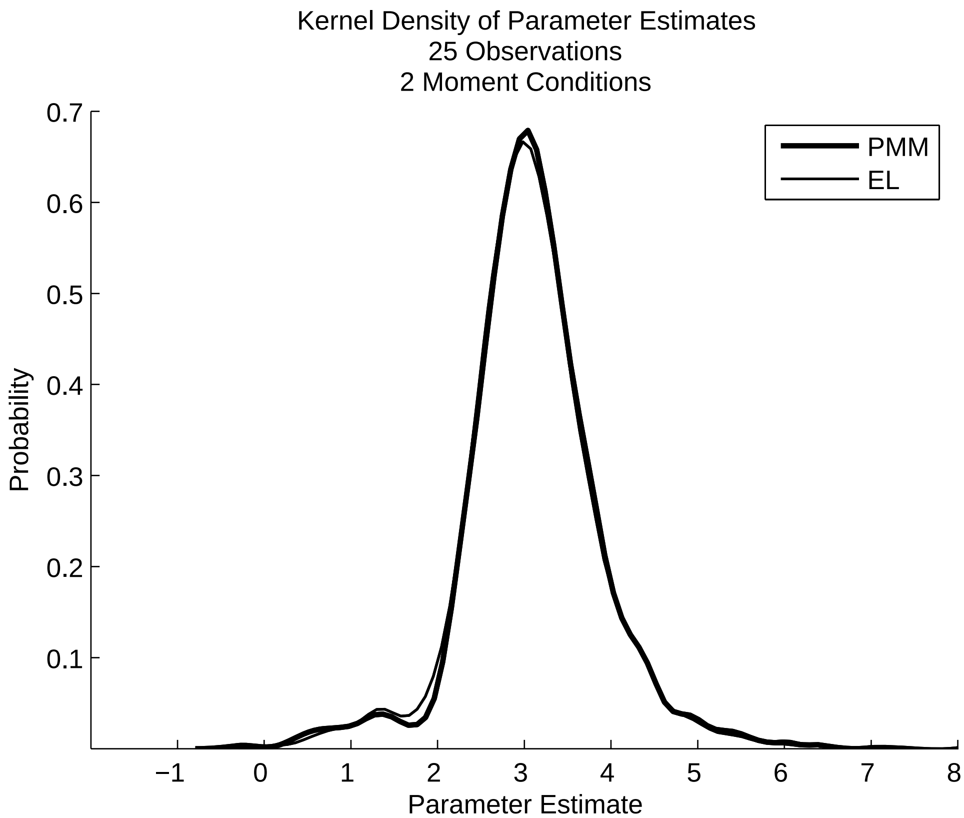

Figure 2: **Kernel Density Plot for 25 Observations and 5 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 25 observations, 5 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

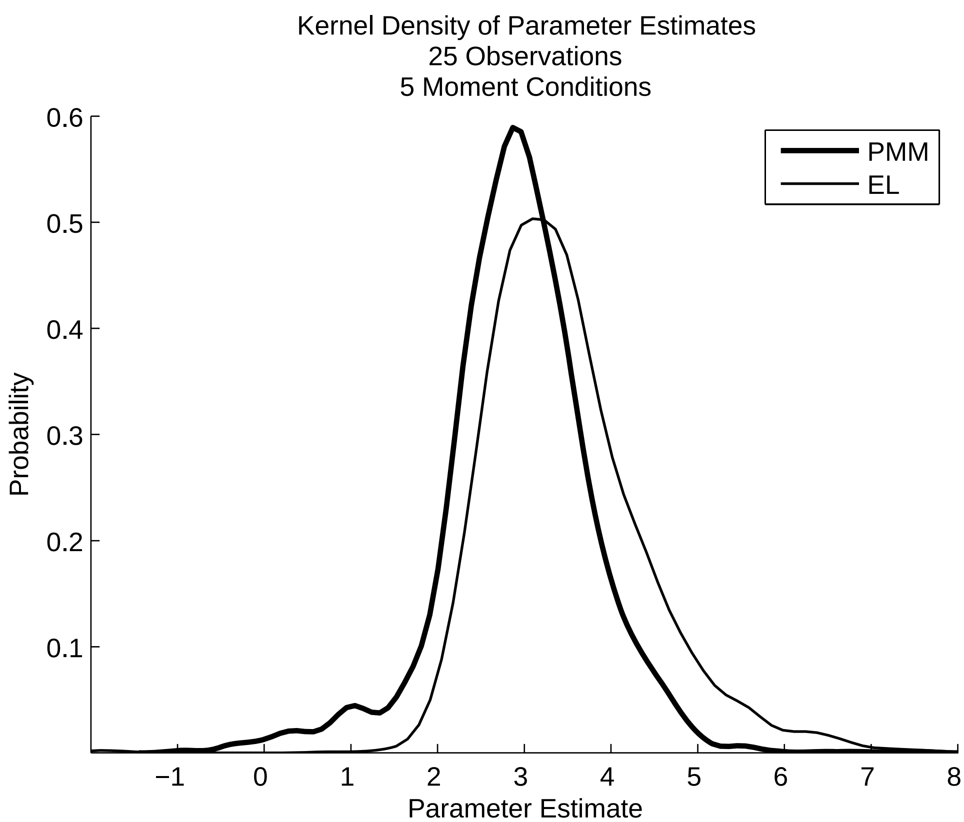

Figure 3: **Kernel Density Plot for 50 Observations and 2 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 25 observations, 2 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

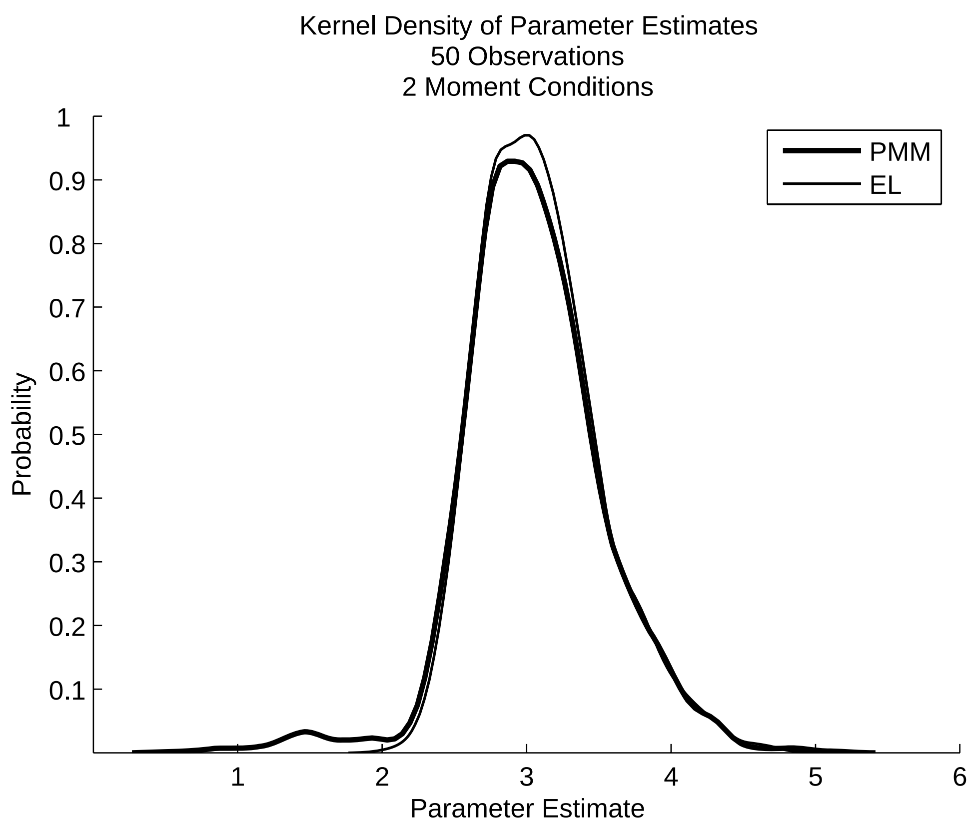

Figure 4: **Kernel Density Plot for 50 Observations and 5 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 50 observations, 5 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

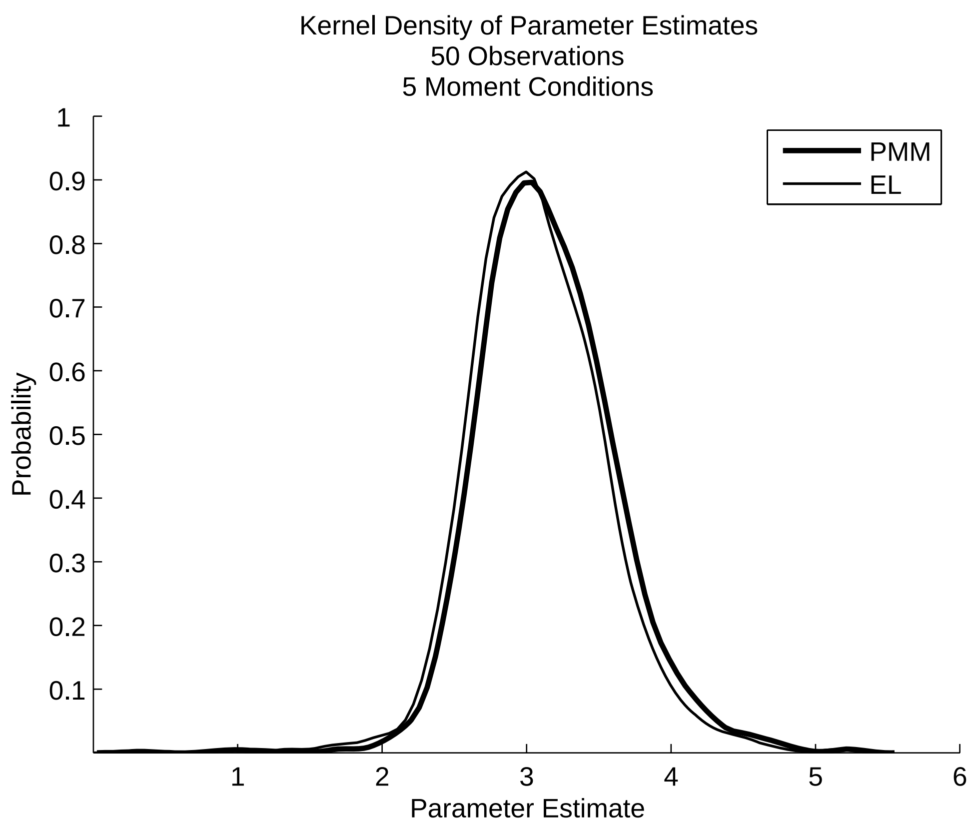

Figure 5: **Kernel Density Plot for 50 Observations and 10 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 50 observations, 10 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

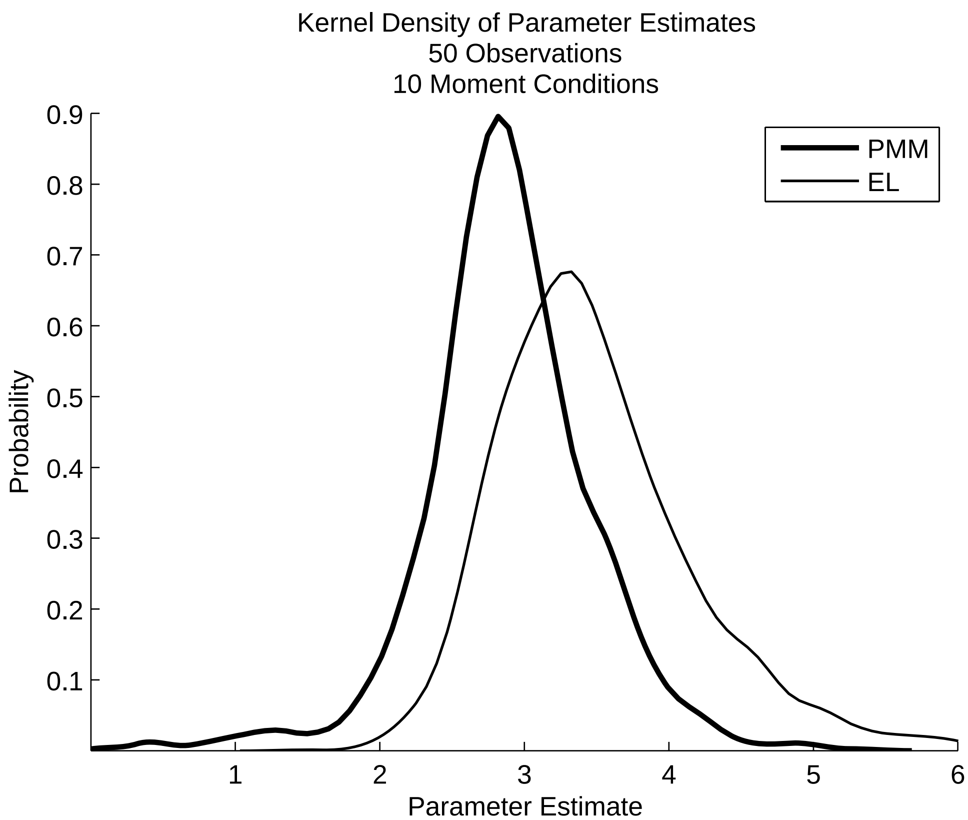

Figure 6: **Kernel Density Plot for 100 Observations and 2 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 100 observations, 2 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

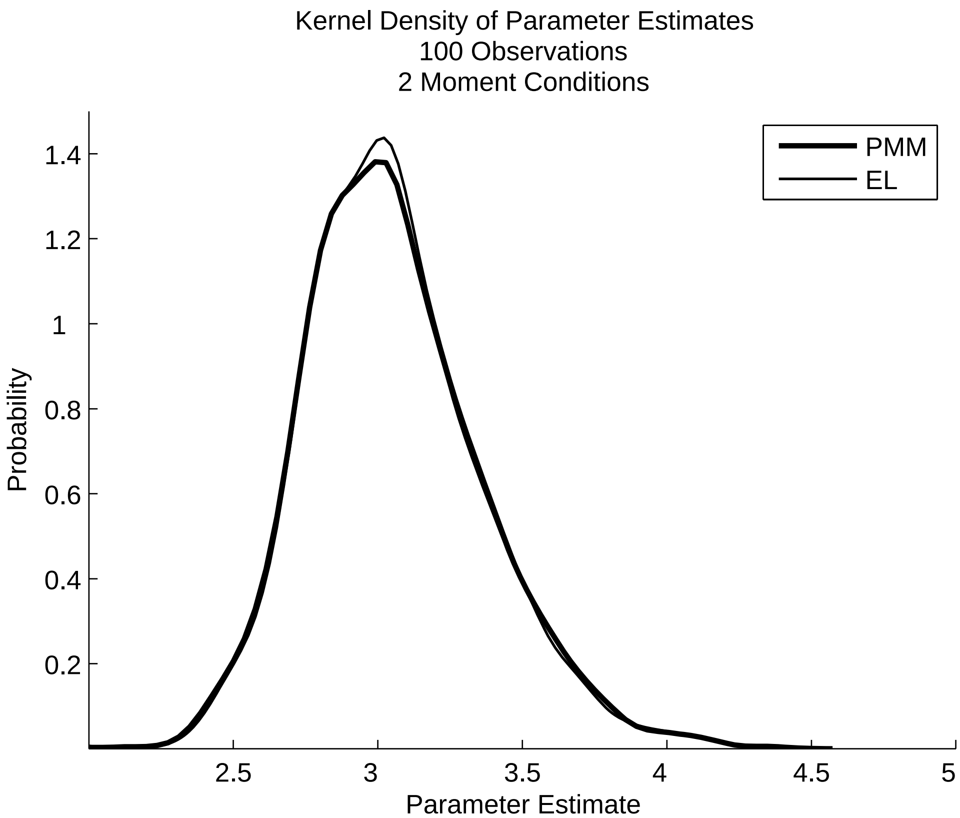

Figure 7: **Kernel Density Plot for 100 Observations and 5 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 100 observations, 5 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

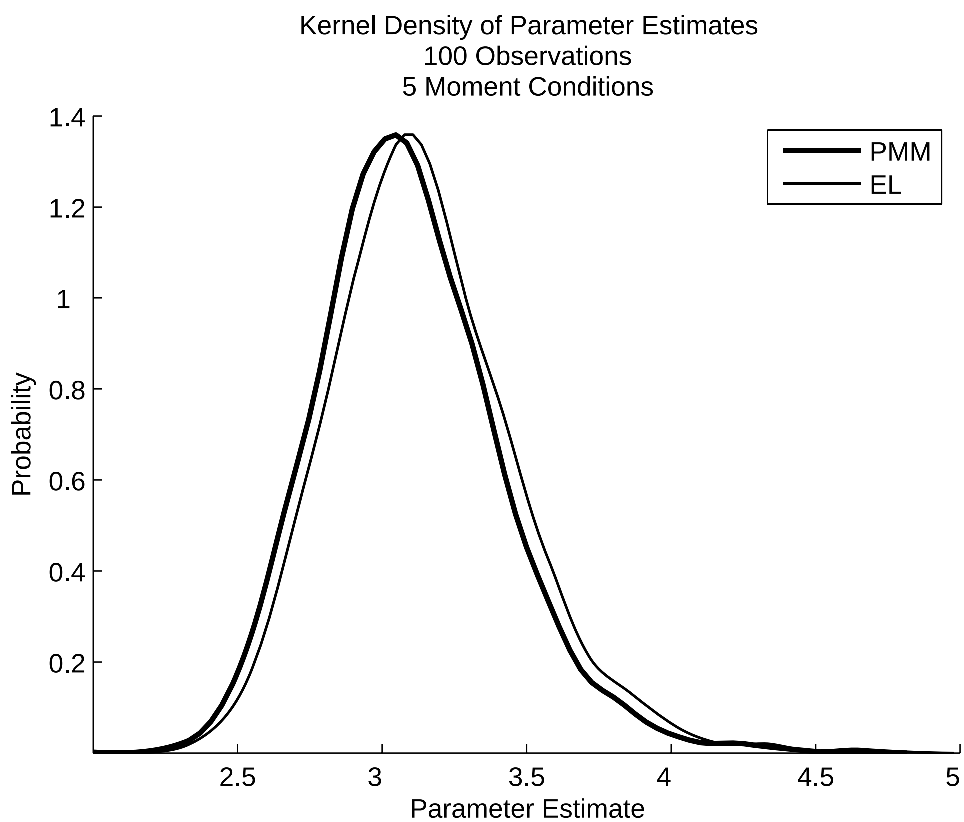

Figure 8: **Kernel Density Plot for 100 Observations and 10 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 100 observations, 10 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

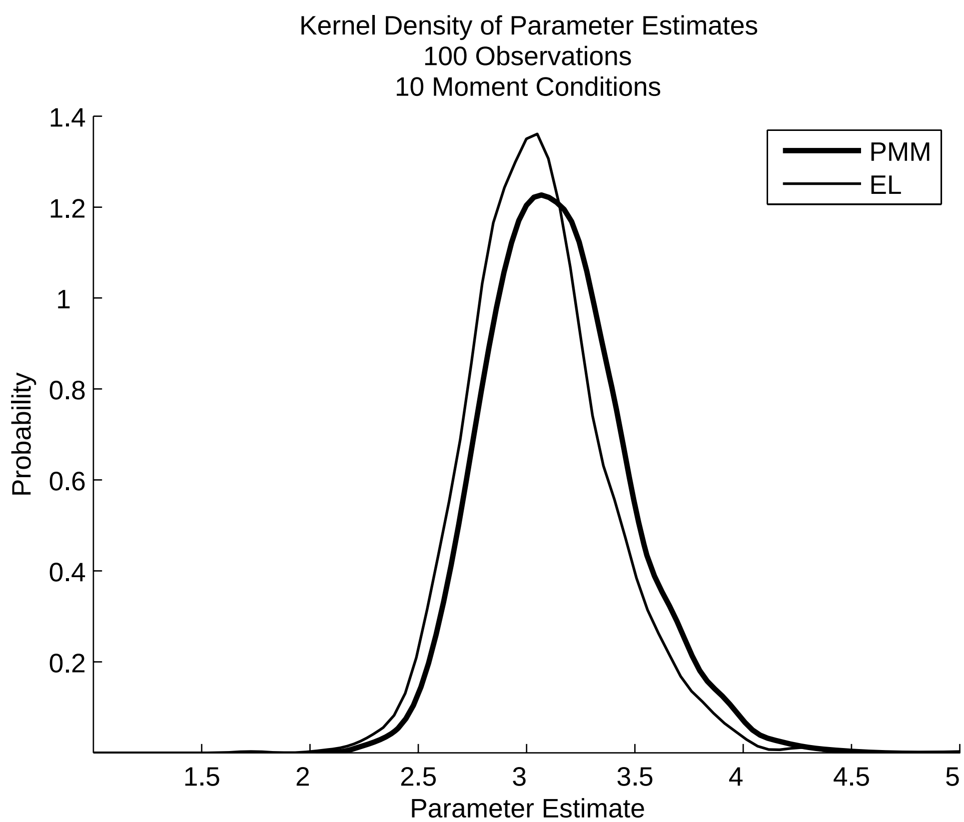

Figure 9: **Kernel Density Plot for 100 Observations and 15 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 100 observations, 15 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.

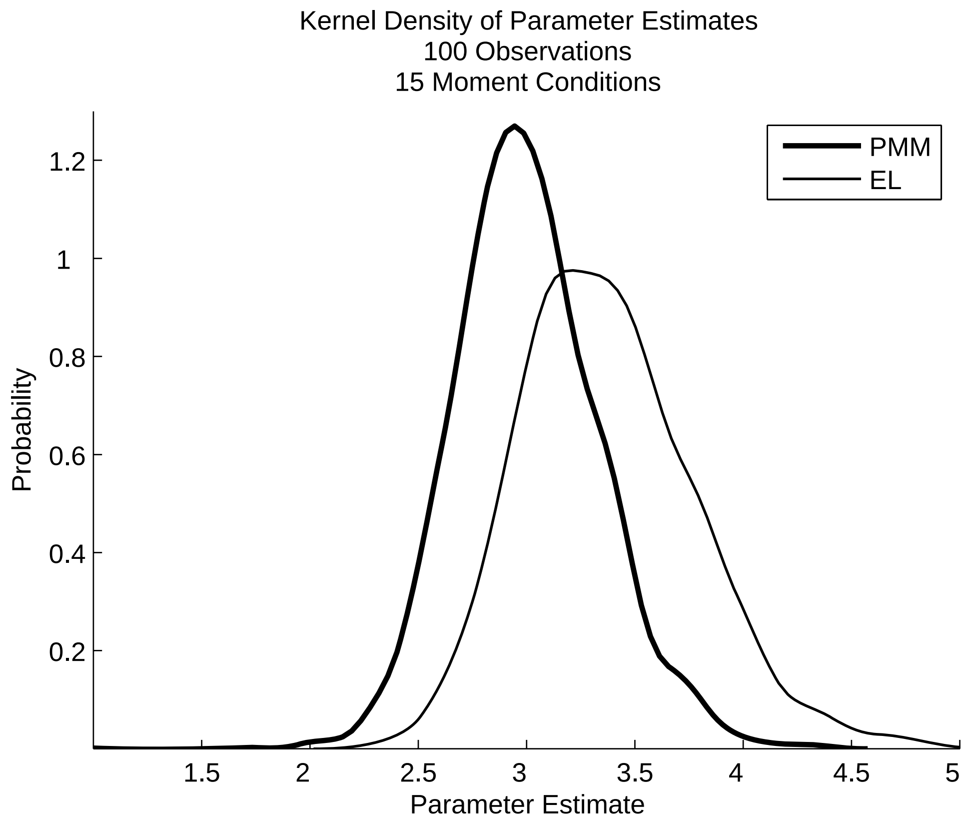

Figure 10: **Kernel Density Plot for 100 Observations and 20 Moment Conditions.** This figure plots the kernel density of the EL and PMM estimates obtained for a Monte Carlo study for the Hall and Horowitz (1996) model as modified by Schennach (2006) with sample size of 100 observations, 20 moment conditions, and 2000 trials. The population parameter value is $\theta_0 = 3$. PMM is represented by the bold line.