Supplement to "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited" - More on Inference*

Joshua D. Angrist  Óscar Jordà  Guido M. Kuersteiner
MIT and NBER  Federal Reserve Bank of San Francisco  University of Maryland
and U.C. Davis

June 12, 2016

1 Setup

This supplement derives the limiting distribution of the specification tests in Appendix D and contains a more detailed discussion of the regularity conditions. We demonstrate that our regularity conditions imply the conditions in Newey and West (1994), justifying the use of their robust standard error estimator.

For ease of reference we repeat a number of definitions from the main paper. The identification restriction is:

**Condition 1** Selection on observables:

\[ y_{t,l}^\psi (d_j) \perp D_t | z_t \text{ for all } l \geq 0 \text{ and for all } d_j, \text{ with } \psi \text{ fixed; } \psi \in \Psi. \]

Let

\[ \delta_{t,j} (\psi) = \delta_{t,j} (z_t, \psi) = \frac{1 \{ D_t = d_j \}}{p^j (z_t, \psi)} - \frac{1 \{ D_t = d_0 \}}{p^0 (z_t, \psi)}, \]

and define the residual weights as \( \tilde{\delta}_{t,j} = \delta_{t,j} (\tilde{\psi}) - \tilde{\delta}_{t,j} \) where \( \tilde{\delta}_{t,j} \) is the predicted value formed from a regression of \( \delta_{t,j} (\tilde{\psi}) \) on \( z_t \), the variables included in the propensity score model. Define \( \hat{h}_{j,t} = Y_{t,L}\hat{\delta}_{t,j} \) and hence \( \hat{h}_t = (\hat{h}_{1,t}, ..., \hat{h}_{J,t})' \). Therefore,

\[ \hat{\theta} = T^{-1} \sum_{t=1}^T \hat{h}_t. \]

---

*Thanks go to Jordi Galí, Ivan Werning, and seminar participants at the Bank of England, the European Central Bank, the Federal Reserve Bank of San Francisco, the NBER Summer Institute 2013, and Yale University for helpful suggestions and comments. The views expressed herein are solely the responsibility of the authors and should not be interpreted as representing the views of the Federal Reserve Bank of San Francisco or the Board of Governors of the Federal Reserve System. Early Elias, Helen Irvin, and Daniel Molitor provided expert research assistance. Contact Information: angrist@mit.edu; oscar.jorda@frb.org and ojorda@ucdavis.edu; kuersteiner@econ.umd.edu*
The estimator $\hat{\theta}$ can also be obtained as the solution to the following minimum distance problem:

$$
\hat{\theta} = \arg \min_{\theta} \left( T^{-1} \sum_{t=1}^{T} \hat{h}_t - \theta \right)' \Omega^{-1} \left( T^{-1} \sum_{t=1}^{T} \hat{h}_t - \theta \right),
$$

(2)

Below we discuss estimates of the spectral density of $\hat{h}_t$ that take into account first step estimation of $\psi$. First note that our estimates of the optimal $\Omega$ are equivalent to estimates of the optimal weight matrix given in Hansen (2008, Section 4.2).

Assume $\hat{\psi}$ is the maximum likelihood estimator with representation

$$
T^{1/2} \left( \hat{\psi} - \psi \right) = \Omega_{\psi}^{-1/2} \sum_{t=1}^{T} l(D_t, z_t, \psi_0) + o_p(1)
$$

(3)

where $\Omega_{\psi} = E\left[ l(D_t, z_t, \psi_0) l(D_t, z_t, \psi_0)' \right]$ and the function

$$
l(D_t, z_t, \psi) = \sum_{j=0}^{J} 1 \{ D_t = d_j \} \frac{\partial p_{t,d_j} (z_t, \psi)}{\partial \psi}
$$

is the score of the maximum likelihood estimator. Define the population projection $\pi_y$ as

$$
\pi_y = \arg \min_{b} E \left[ \| Y_{t,L} - b z_t \| ^2 \right],
$$

define $\vartheta = (\psi', (\text{vec } \pi_y)')'$ and let $h_t (\vartheta_0) = (Y_{t,L} - \pi_y z_t) \delta_{t,j} (\psi_0)$. The representation in (3) is used to expand $\hat{h}_t$ around $\psi_0$ leading to $\hat{\theta} - \theta_0 = T^{-1} \sum_{t=1}^{T} v_t (\vartheta_0) + o_p \left( T^{-1/2} \right)$ where $v_t (\vartheta_0) = h_t (\vartheta_0) - \theta_0 + h(\vartheta_0) \Omega_{\psi}^{-1} l(D_t, z_t, \psi_0)$ and $h(\vartheta_0) = E \left[ \partial h_t (\vartheta_0) / \partial \vartheta' \right]$. The covariance matrix $\Omega_{\theta}$ is the typical spectrum at frequency zero matrix of $v_t (\vartheta_0)$ found in the HAC-standard error literature (see Newey and West (1994)) and is given by

$$
\Omega_{\theta} = \sum_{i=-\infty}^{\infty} E \left[ v_t (\vartheta_0) v_{t-i} (\vartheta_0)' \right]
$$

(4)

The formula for $\Omega_{\theta}$ takes into account that the ‘observations’ $\hat{h}_t$ used to compute the sample averages are based on estimated, rather than observed data. Confidence intervals for $\theta$ can be constructed from $\Omega_{\theta}$. We use the procedure in Newey and West (1994) to estimate $\Omega_{\theta}$. Below, we provide further details regarding regularity conditions needed for the Newey West procedure.

Using $\hat{\vartheta} = (\hat{\psi}', (\text{vec } \pi_y)')'$ where $\hat{\pi}_y$ is the OLS estimator in a regression of $Y_{t,L}$ on $z_t$ we estimate $\Omega_{\theta}$ from the sample averages

$$
\hat{h}(\vartheta) = T^{-1} \sum_{t=1}^{T} \partial h_t \left( \hat{\vartheta} \right) / \partial \vartheta', \quad \hat{\Omega}_{\psi} = -T^{-1} \sum_{t=1}^{T} \partial l(D_t, z_t, \hat{\psi}) / \partial \vartheta'
$$

and by letting $v_t \left( \hat{\vartheta} \right) = h_t \left( \hat{\vartheta} \right) - \hat{\theta} + \hat{h}(\hat{\vartheta}) \hat{\Omega}_{\psi}^{-1} l(D_t, z_t, \hat{\psi})$. As in Newey and West (1994), we use the Bartlett
kernel with prewhitening and a data-dependent plug in estimator to obtain the necessary bandwidth parameter.

The Newey and West procedure is implemented as follows. Prewhitening is achieved by fitting a AR(1) model to each element $v_{t,j} \left( \hat{\vartheta} \right)$ of $v_t \left( \hat{\vartheta} \right)$. For this purpose define the autoregressive parameter estimate

$$
\hat{A}_{jj} = \sum_{t=2}^{T} v_{t,j} \left( \hat{\vartheta} \right) v_{t-1,j} \left( \hat{\vartheta} \right)^\prime \left( \sum_{t=2}^{T} v_{t-1,j} \left( \hat{\vartheta} \right) v_{t-1,j} \left( \hat{\vartheta} \right)^\prime \right)^{-1}
$$

and let $\hat{r}_t \left( \hat{\vartheta} \right) = v_t \left( \hat{\vartheta} \right) - \hat{A} v_{t-1} \left( \hat{\vartheta} \right)$ where $\hat{A}$ is a diagonal matrix with diagonal elements $\hat{A}_{jj}$. Then define $\hat{\Omega}_{\Theta,j} = (T-1)^{-1} \sum_{t=j+1}^{T} \hat{r}_t \left( \hat{\vartheta} \right) \hat{r}_{t-j} \left( \hat{\vartheta} \right)^\prime$ for $j \geq 0$ and $\hat{\Omega}_{\Theta,j} = \hat{\Omega}_{\Theta,-j}$ for $j < 0$. Let $1 = [1, \ldots, 1]^\prime$ be an $r$-dimensional vector where $r$ is the dimension of $\theta$. Define $\tilde{\delta}_j = 1 \hat{\Omega}_{\Theta,j} 1$, $\tilde{\gamma} = \sqrt{\sum_{j=-n}^{n} \left( j \right)^3 \tilde{\gamma}_j}$ and $\hat{\gamma} = c_{\gamma} \left( \tilde{\gamma}/\tilde{\gamma} \right)^{2/3}$ where $c_{\gamma} = 1.1447$ and $n = \left[ 4 \left( T/100 \right)^2 / 9 \right]$ where $\left[ . \right]$ denotes the integer part of a real number. Set the bandwidth parameter to $\tilde{B} = \left\lceil \hat{\gamma} T^{1/3} \right\rceil$.

The estimator for $\hat{\Omega}_{\Theta}$ is now defined as

$$
\hat{\Omega}_{\Theta} = \left( I_r - \hat{A} \right)^{-1} \left( \Omega_{\Theta,0} + \sum_{j=1}^{\tilde{B}} \left( 1 - \frac{j}{\tilde{B} + 1} \right) \left( \hat{\Omega}_{\Theta,j} + \hat{\Omega}_{\Theta,-j} \right) \right) \left( I_r - \hat{A} \right)^{-1}.
$$

An important diagnostic for our purposes looks at whether lagged macro aggregates are independent of policy changes conditional on the policy propensity score. In other words we would like to show that the policy shocks implicitly defined by our score model look to be “as good as randomly assigned.” Angrist and Kuersteiner (2011) develop semiparametric tests that can be used for this purpose.

The specification tests are based on the following fact. If $w_t$ is a vector of $k_w$ elements of $z_t$ or $\chi_{t-1}$, then correct specification of the propensity score implies that

$$
E \left[ \delta_{t,j} (\psi_0) | w_t \right] = 0 \text{ for all } j = 1, \ldots, J.
$$

All $J$ conditional moment restrictions, or a subset of them, can be summarized into a vector. Let $D_t (z_t, \psi) = (\delta_{t,j_1} (\psi), \ldots, \delta_{t,j_k} (\psi))$. Set $k \leq J$ and $1 \leq j_1 < \ldots < j_k \leq J$. In our case, we use this setup to focus on $d_j = \{-.25, 0, .25\}$. Then, $E \left[ D_t (z_t, \psi_0) | w_t \right] = 0$ must hold. To test this condition, consider the unconditional moment restriction $E \left[ D_t (z_t, \psi_0) \otimes w_t \right] = 0$. Since our estimators are based on $\hat{\delta}_{t,j}$ we similarly define our test based on $\hat{\delta}_{t,j}$. For this purpose, let $\tilde{D}_t (z_t, \psi) = (\hat{\delta}_{t,j_1} (\psi), \ldots, \hat{\delta}_{t,j_k} (\psi))$ and consider the test statistic $T^{-1/2} \sum_{t=1}^{T} \tilde{D}_t (z_t, \psi) \otimes w_t$. Let $\pi_w$ be the population projection parameter of a projection of $w_t$ onto $z_t$, and $\hat{\pi}_w$ the corresponding sample OLS estimator. Define $\xi = (\psi', \text{vec} (\pi_w))'$, let $m_t (\xi) = (D_t (z_t, \psi)) \otimes (w_t - \pi_w z_t)$ and define $\tilde{m} (\xi) = T^{-1} \sum_{t=1}^{T} m_t (\xi)$. It then follows that $T^{-1} \sum_{t=1}^{T} \tilde{D}_t (z_t, \psi) \otimes w_t = \tilde{m} (\hat{\xi})$ where $\hat{\xi} = (\hat{\psi}', \text{vec} (\hat{\pi}_w))'$ and we base our statistic on $\tilde{m} (\hat{\xi})$. The limiting distribution of $\tilde{m} (\hat{\xi})$ is affected by the fact that $\psi_0$ is estimated. Define $\tilde{m} (\xi) = E \left[ \partial m_t (\xi) / \partial \psi' \right],

---

1See Newey and West (1994, Tables I and II).
\( \hat{m}_t = m_t \left( \tilde{\xi} \right) \) and consider the expansion

\[
\hat{m}_t = m_t (\xi_0) + \hat{m} (\xi_0) \Omega_{\psi}^{-1} l(D_t, z_t, \psi_0) + o_p \left( T^{-1/2} \right).
\]

A key insight is that under the null-hypothesis, \( \hat{m}_t \) is approximately a martingale difference sequence and thus is mean zero. This feature significantly simplifies estimation of the asymptotic variance normalizing the test. Then, letting \( \hat{m} = \hat{m} \left( \tilde{\xi} \right) \), \( \nu_t (\xi_0) = m_t (\xi_0) + \hat{m} (\xi_0) \Omega_{\psi}^{-1} l(D_t, z_t, \psi_0) \) and \( \hat{V} = T^{-1} \sum_{t=1}^{T} \nu_t \left( \tilde{\xi} \right) \nu_t \left( \tilde{\xi} \right)' \) leads to the test statistic

\[
T \hat{m}' \hat{V}^{-1} \hat{m} \rightarrow_d \chi^2_{(k-k_0)}
\]

under the null hypothesis that \( E \left[ 1 \{ D_t = j \} | z_t \right] = p^j (z_t, \psi_0) \). The limiting distribution in (5) is established below.

## 2 Regularity Conditions

We repeat some of the definitions and derivations already reported in the paper to make the supplement easier to follow. Assume that \( \{ \chi_t \}_{t=-\infty}^{\infty} \) is strictly stationary with values in the measurable space \((\mathbb{R}^r, \mathcal{B}^r)\) where \( \mathcal{B}^r \) is the Borel field on \( \mathbb{R}^r \) and \( r \) is fixed with \( 2 \leq r < \infty \). Let \( \mathcal{A}_k = \sigma (\chi_k, \ldots, \chi_t) \) be the sigma field generated by \( \chi_k, \ldots, \chi_t \). The sequence \( \chi_t \) is \( \varphi \)-mixing if

\[
\varphi_m = \sup_l \left[ \sup_{A \in \mathcal{A}_\infty^{\infty}, B \in \mathcal{A}_l, P(B) > 0} |Pr (A | B) - P (A) | \right] \rightarrow 0 \text{ as } m \rightarrow \infty.
\]

**Condition 2** Let \( \chi_t \) be a stationary, \( \varphi \)-mixing sequence such that for some \( 2 < p < \infty \) the \( \varphi \)-mixing coefficient of \( \chi_t \) satisfies \( \varphi_m \leq cm^{-1} m^{-1/p} \) for some bounded constant \( c > 0 \). For each element \( \chi_{t,j} \) of \( \chi_t \) it follows that \( E \left[ |\chi_{t,j}|^p \right] < \infty \).

Condition 2 implies that \( \sum_{m=1}^{\infty} \varphi_m^{1-1/p} < \infty \) as required for Corollary 3.9 of McLeish (1975a). In addition, \( \varphi_m \) satisfies (2.6) of McLeish (1975b) required for a strong law of large numbers. This follows because for any \( p > 2 \) the inequality \( p/(p - 2) < (1 + p) / (p - 4/p) \) holds and, since \( p > 2 \), the moment restrictions imposed below are stronger than required by McLeish. Using Corollary A.2 of Hall and Heyde (1980), and assuming that, for each element \( v_{t,j} (\vartheta_0) \) of \( v_t (\vartheta_0) \), \( E \left[ |v_{t,j} (\vartheta_0)|^p \right] < \infty \) it also follows that \( \sum_{m=1}^{\infty} m^q \| E \left[ v_t (\vartheta_0) v_t - m (\vartheta_0)' \right] \| < \infty \) for some \( q > 7/4 \) as required by Assumption 2 of Newey and West (1994) when the Bartlett kernel is used. If the size of the mixing coefficients is weakened to \( - (1 + p) / (p - 2/p) \) then Assumption 2 of Newey and West holds for all \( p > 2 + \sqrt{6} \) and some \( q > 7/4 \). Also note that \( p > 2 \) is sufficient to satisfy Assumption 3 of Newey and West (1994) when the Bartlett kernel is used as suggested here.
The next condition states that the propensity score \( p(z_t, \theta) \) is the correct parametric model for the conditional expectation of \( D_t \) and lists a number of additional regularity conditions.

**Condition 3** Let \( \Theta \) be a compact subset of \( \mathbb{R}^{k_\theta} \) where \( k_\theta \) is the dimension of \( \theta \). Let \( \psi_0 \in \Psi \subset \Theta \) where \( \Psi \subset \mathbb{R}^{k_\psi} \) is a compact set and \( k_\psi < \infty \). Assume that \( E \left[ \{ D_t = d_j \} | z_t \right] = p_t^j (z_t, \psi_0) \) and for all \( \psi \neq \psi_0 \) it follows \( E \left[ \{ D_t = d_j \} | z_t \right] = p_t^j (z_t | \psi) \). Assume that \( p_t^j (z_t | \psi) \) is differentiable a.s. for \( \theta \in \{ \theta \in \Theta \mid \| \theta - \hat{\theta}_0 \| \leq \delta \} := N_\delta (\theta_0) \) for some \( \delta > 0 \). Let \( N(\theta_0) \) be a compact subset of the union of all neighborhoods \( N_\delta (\theta_0) \) where \( \partial p_t^j (z_t | \psi) / \partial \psi \), \( \partial^2 p_t^j (z_t | \psi) / \partial \psi \partial \psi \) exists and assume that \( N(\theta_0) \) is not empty. Assume that for all \( j \in \{ 0, \ldots, J \} \) and some \( \delta_0 > 0 \) and any \( \delta > 0 \), \( \theta, \theta^* \) with \( \| \theta - \theta^* \| < \delta \leq \delta_0 \) there exists a random variable \( B_t \) which is a measurable function of \( D_t, z_t \) and \( Y_{t,L} \) and a constant \( \alpha > 0 \) such that for all \( i \)

\[
\| h_{t,j} (\theta) - h_{t,j} (\theta^*) \| \leq B_t \| \theta - \theta^* \|^\alpha,
\]

and

\[
\| \partial h_{t,j} (\theta) / \partial \theta - \partial h_{t,j} (\theta^*) / \partial \theta \| \leq B_t \| \theta - \theta^* \|^\alpha, \tag{6}
\]

\[
\| \partial^2 h_{t,j} (\theta) / \partial \theta \partial \theta' - \partial^2 h_{t,j} (\theta^*) / \partial \theta \partial \theta' \| \leq B_t \| \theta - \theta^* \|^\alpha, \tag{7}
\]

\[
\| z_t (\delta_{t,j} (\psi) - \delta_{t,j} (\psi^*)) \| \leq B_t \| \psi - \psi^* \|^\alpha. \tag{8}
\]

and \( \theta, \theta^* \in \text{int}(N(\theta_0)) \). Let \( h_{t,j,i} (\theta) \) be the \( i \)-th element of \( h_{t,j} (\theta) \) and \( \hat{\theta}_k \) the \( k \)-th element of \( \theta \). Assume \( E \| B_t \|^p \) \( < \infty \), and for all \( i, j \) that \( E \| h_{t,j,i} (\theta_0) \|^p \) \( < \infty \), \( E \| \partial h_{t,j,i} (\theta_0) / \partial \theta \partial \theta' \|^p \) \( < \infty \), and

\[
E \left[ \| \partial^2 h_{t,j,i} (\theta_0) / \partial \theta \partial \theta' \|^p \right] \leq \infty.
\]

**Condition 4** Assume that \( \hat{\theta} - \theta_0 = \alpha_p (1), T^{1/2} \left( \hat{\psi} - \psi_0 \right) = \Omega_p^{-1} T^{-1/2} \sum_{t=1}^T l(D_t, z_t, \psi_0) + \alpha_p (1) \). Assume that \( E [z_t z_t'] \) is positive definite. Let \( l_t (D_t, z_t, \psi_0) \) be the \( i \)-th element of \( l(D_t, z_t, \psi) \). Let \( p \) be given as in Condition 2 and assume that \( E \| l(D_t, z_t, \psi_0) \|^p < \infty \), \( \sup_{\psi \in N(\theta_0)} \| l(D_t, z_t, \psi) \| \leq B_t \),

\[
\sup_{\psi \in N(\theta_0)} \| \partial l(D_t, z_t, \psi) / \partial \psi \| \leq B_t
\]

and \( \sup_{\psi \in N(\theta_0)} \| \partial^2 l_t (D_t, z_t, \psi) / \partial \psi \partial \psi \| \leq B_t \).

**Condition 5** Assume that \( \Omega_p \) is positive definite for all \( \psi \) in some neighborhood \( N \subset \Psi \) such that \( \psi_0 \in \text{int} N \) and \( 0 < \| \Omega_p \| < \infty \) for all \( \psi \in N \). Assume that \( \Omega_\theta \) defined in (4) is positive definite.

Conditions 2, 3 and 4 imply that Assumption 2 of Newey and West is satisfied. The results of their paper thus apply to the estimates of \( \Omega_\theta \) proposed here.

Regularity conditions for the specification tests are given below.
Condition 6 Let $N(\xi_0)$ a neighborhood of $\xi_0$ defined similarly to the one in Condition 3. Let $p$ be given as in Condition 2. For some random variable $B_t$ which is a measurable function of $D_t, z_t$ and $w_t$ and for which $E[B_t^p] < \infty$, it holds that for some $\varepsilon > 0$ and $\xi, \xi^*$ with $\|\xi - \xi^*\| < \varepsilon \leq \delta_0$ and $\xi, \xi^* \in \text{int} N(\xi_0)$ that

i) $E \left[ m_t(\xi_0)^{\varepsilon} \right] < \infty$, $E \left[ |\partial m_t(\xi_0)| / \partial z^\varepsilon \right] < \infty$, $E \left[ |l(D_t, z, \psi_0)| \right] < \infty$

ii) $|l(D_t, z, \psi) - l(D_t, z, \psi^*)| \leq B_t \|\psi - \psi^*\|^\alpha$,

iii) $|\partial m_t(\xi) / \partial z' - \partial m_t(\xi^*) / \partial z'| \leq B_t \|\xi - \xi^*\|^\alpha$.

3 Proofs

The proof of the following theorem appears in the Appendix to the paper and is repeated here for convenience.

Theorem 1 Let $\hat{\theta}$ be defined in (1) and assume that Conditions 1, 2, 3, 4, and 5 hold. Then, $\hat{\theta} \to_p \theta$ and

$$T^{1/2} \left( \hat{\theta} - \theta \right) \overset{d}{\to} N(0, \Omega_0)$$

where $\Omega_0$ is defined in (4).

Proof. Let $Z = (z_1, ..., z_T)'$, $Y_L = (Y_{1,L}, ..., Y_{T,L})'$ and $\delta_j \left( \hat{\psi} \right) = (\hat{\delta}_{1,j}(\hat{\psi}), ..., \hat{\delta}_{T,j}(\hat{\psi}))'$. Define the population projection $\pi_y$ as $\pi_y = \arg \min_{\pi_y} E \left[ \|Y_{L,L} - h\pi_L\|^2 \right]$ and sample analog $\hat{\pi}_y = Y_L'Z (Z'Z)^{-1}$. Recall that $\hat{h}_{t,j} = Y_{t,L} \left( \hat{\delta}_{t,j}(\hat{\psi}) - \hat{\delta}_{t,j} \right)$ where $\hat{\delta}_{t,j} = z_t'(Z'Z)^{-1}Z'\delta_j(\hat{\psi})$ and let $h_{t,j}(\theta_0) = (Y_{t,L} - \pi_y z_t) \delta_{t,j}(\psi_0)$.

First observe that

$$\sum_{t=1}^T \hat{h}_{t,j} = \sum_{t=1}^T Y_{t,L} \left( \delta_{t,j}(\hat{\psi}) - \hat{\delta}_{t,j} \right) = \sum_{t=1}^T Y_{t,L} \delta_{t,j}(\hat{\psi}) - \sum_{t=1}^T Y_{t,L} z_t'(Z'Z)^{-1} \sum_{s=1}^T z_s' \delta_{s,j}(\hat{\psi}) = \sum_{t=1}^T Y_{t,L} - \hat{\pi}_y z_t' \delta_{t,j}(\hat{\psi}).$$

By the Mean Value Theorem we then obtain

$$T^{1/2} \left( \hat{\theta}_j - \theta_{0,j} \right) = T^{-1/2} \sum_{t=1}^T \hat{h}_{t,j} - \theta_{0,j}$$

$$= T^{-1/2} \sum_{t=1}^T (Y_{t,L} - \pi_y z_t) \delta_{t,j}(\hat{\psi}) - \theta_0 + (\pi_y - \hat{\pi}_y) T^{-1/2} \sum_{t=1}^T z_t \delta_{t,j}(\hat{\psi})$$

$$= T^{-1/2} \sum_{t=1}^T h_{t,j}(\theta_0) - \theta_0 + T^{-1} \sum_{t=1}^T \partial h_{t,j}(\theta_0) / \partial \psi' T^{1/2} (\hat{\psi} - \psi_0)$$

$$+ T^{-1} \sum_{t=1}^T (\partial h_{t,j}(\hat{\theta}) / \partial \psi' - \partial h_{t,j}(\theta_0) / \partial \psi') T^{1/2} (\hat{\psi} - \psi_0) + (\pi_y - \hat{\pi}_y) T^{-1/2} \sum_{t=1}^T z_t \delta_{t,j}(\hat{\psi}).$$
where $\|\hat{\psi} - \vartheta_0\| \leq \|\hat{\vartheta} - \vartheta_0\|$ and $\partial h_t(\vartheta) / \partial \psi' = [\partial h_{t,1}(\vartheta) / \partial \psi', \ldots, \partial h_{t,j}(\vartheta) / \partial \psi']$ with

$$\partial h_{t,j}(\vartheta) / \partial \psi = (Y_{t,L} - \pi_y z_t) \left( - \frac{D_{t,j}}{\tilde{p}^j(z_t, \psi)} \frac{\partial \tilde{p}^j(z_t, \psi)}{\partial \psi} + \frac{D_{t,0}}{p^0(z_t, \psi)} \frac{\partial \rho^0(z_t, \psi)}{\partial \psi} \right). \quad (10)$$

By (6) it follows that for $\delta_0$ given in Condition 3 and any $\delta$ such that $\delta_0 > \delta > 0$,

$$P \left( \left\| T^{-1} \sum_{t=1}^{T} (\partial h_{t,j}(\vartheta) / \partial \psi' - \partial h_{t,j}(\vartheta_0) / \partial \psi') \right\| > \eta \right) \leq P \left( \sup_{\|\vartheta - \vartheta_0\| \leq \delta} \left\| T^{-1} \sum_{t=1}^{T} (\partial h_{t,j}(\vartheta) / \partial \psi' - \partial h_{t,j}(\vartheta_0) / \partial \psi') \right\| > \eta, \|\hat{\vartheta} - \vartheta_0\| < \delta \right) + P \left( \|\hat{\vartheta} - \vartheta_0\| \geq \delta \right)$$

$$= \frac{E[|B_t|^p]}{\eta^p} \delta^p + P \left( \|\hat{\vartheta} - \vartheta_0\| \geq \delta \right)$$

where both terms can be made arbitrarily small by choosing $\eta = \sqrt{\delta}$ and $\delta > 0$ for $T$ large enough by using Conditions 4 and 3. By McLeish (1975b, Theorem 2.10) $T^{-1} \sum_{t=1}^{T} \partial h_{t,j}(\vartheta) / \partial \psi' \xrightarrow{p} \dot{h}_j(\vartheta_0)$ where we defined $E[\partial h_{t,j}(\vartheta_0) / \partial \psi'] = \dot{h}_j(\vartheta_0)$. This implies that the third term in (9) is $o_p(1)$.

For the last term in (9) note that $(\pi_y - \hat{\pi}_y) = O_p \left( T^{-1/2} \right)$ by McLeish (1975b, Theorem 2.10), Corollary 3.9 of McLeish (1975a) and standard arguments for linear regressions. Now consider

$$\begin{align*}
(\pi_y - \hat{\pi}_y) T^{-1/2} \sum_{t=1}^{T} z_t \delta_{t,j}(\psi) \\
= T^{1/2} (\pi_y - \hat{\pi}_y) T^{-1} \sum_{t=1}^{T} z_t \delta_{t,j}(\psi_0) \\
+ T^{1/2} (\pi_y - \hat{\pi}_y) T^{-1} \sum_{t=1}^{T} z_t \left( \delta_{t,j}(\psi) - \delta_{t,j}(\psi_0) \right).
\end{align*} \quad (12)$$

The first term in (12) is $o_p(1)$ because from $E[z_t \delta_{t,j}(\psi_0)] = 0$ it follows that

$$T^{-1} \sum_{t=1}^{T} z_t \delta_{t,j}(\psi_0) = o_p(1). \quad (13)$$

For the second term in (12) use Condition 3 to show that

$$T^{-1} \sum_{t=1}^{T} z_t \left( \delta_{t,j}(\psi_0) - \delta_{t,j}(\hat{\psi}) \right) = o_p(1) \quad (14)$$

by arguments similar to those in (11). Then, (13) and (14) establish that (12) is $o_p(1)$. It then follows from (12) and (14) that (9) is

$$\begin{align*}
T^{-1/2} \sum_{t=1}^{T} h_{t,j}(\vartheta_0) - \theta_0 \\
+ T^{-1} \sum_{t=1}^{T} \partial h_{t,j}(\vartheta_0) / \partial \psi' T^{1/2} (\hat{\psi} - \psi_0) + o_p(1) \\
= T^{-1/2} \sum_{t=1}^{T} \left[ h_{t,j}(\vartheta_0) - \theta_0 + h_j(\vartheta_0) \Omega_\psi^{-1} l(D_t, z_t, \psi_0) \right] + o_p(1).
\end{align*}$$
Stack \( h_t(\vartheta) = [h_{t,1}(\vartheta)' , \ldots , h_{t,J}(\vartheta)']' \) and \( \hat{h}(\vartheta) = [\hat{h}_1(\vartheta)', \ldots , \hat{h}_J(\vartheta)']' \), let

\[
v_t(\vartheta_0) = h_t(\vartheta_0) - \theta + \hat{h}(\vartheta_0)\Omega^{-1}_\psi l(D_t, z_t, \psi_0)
\]

and \( v_{t,j}(\vartheta_0) \) is the \( j \)-th element of \( v_t(\vartheta_0) \). Note that \( v_{t,j}(\vartheta_0) \) is \( \beta \)-mixing with \( E[v_{t,j}(\vartheta_0)] = 0 \). Then it follows that

\[
T^{-1}E \left[ \sum_{t=1}^T \sum_{s=1}^T v_t(\vartheta_0) v_s(\vartheta_0) \right] = \sum_{j=-T+1}^{T-1} \left( 1 - \frac{|j|}{T} \right) E \left[ v_1(\vartheta_0) v_{1-j}(\vartheta_0) \right] \to \Omega_\vartheta
\]

by stationarity of \( v_t = v_t(\vartheta_0) \) and the Toeplitz lemma. Fix \( \lambda \in \mathbb{R}^k \) with \( \|\lambda\| = 1 \) and let \( S_T = T^{-1/2} \sum_{t=1}^T \lambda' v_t \). Then, \( E[S_T^2] \to \lambda' \Omega_\vartheta \lambda > 0 \) by (15) and Condition 5. In addition

\[
E \left[ \|\lambda' v_t\|^p \right] \leq E \left[ \left( \sum_{l=1}^k |\lambda_l| \|v_{t,l}\| \right)^p \right] \leq \left( \sum_{l=1}^k |\lambda_l| \|v_{t,l}\|^p \right)^{p-1} E \left[ \sum_{l=1}^k |v_{t,l}|^p \right]
\]

by Hölder’s inequality (Magnus and Neudecker, 1988, p.220) and where \( \tilde{v}_{t,i} \) is the \( i \)-th element of \( v_t \). Since \( p/(p-1) \leq 2 \) and \( \|\lambda\| = 1 \) it follows that \( \sum_{l=1}^k |\lambda_l| \|v_{t,l}\| < k \). Denote by \( h_{t,j}(\vartheta_0) \) and \( \theta(j) \) the \( j \)-th element of \( h_t(\vartheta_0) \) and \( \theta \) respectively and by \( \hat{h}_j(\vartheta_0) \) the \( j \)-th row of \( \hat{h}(\vartheta_0) \). Then,

\[
E[|\tilde{v}_{t,j}|^p] \leq E \left[ \left( |h_{t,j}(\vartheta_0)| + |\theta(j)| + \|\hat{h}_j(\vartheta_0)\| \|\Omega^{-1}_\psi\| \|l(D_t, z_t, \psi_0)\| \right)^p \right] \\
\leq 3^{p-1} \left( E[|h_{t,j}(\vartheta_0)|^p] + |\theta(j)|^p + \|\hat{h}_j(\vartheta_0)\|^p \|\Omega^{-1}_\psi\| \|l(D_t, z_t, \psi_0)\|^p \right)
\]

again by Hölder’s inequality. It follows that \( |\theta(j)|^p \leq E[|h_{t,j}(\vartheta_0)|^p] \) by Jensen’s inequality and \( \|\Omega^{-1}_\psi\|^p < \infty \) by Condition 5. Similarly, \( E[\|l(D_t, z_t, \psi_0)\|^p] < \infty \) by Condition 4 and

\[
|\hat{h}_j(\vartheta_0)|^p \leq E[|\partial h_{t,j}(\vartheta_0)/\partial \psi|^p] < \infty
\]

by Condition 3. By Condition 3 \( E[|h_{t,j}(\vartheta_0)|^p] < \infty \) such that \( E[|\tilde{v}_{t,j}|^p] < \infty \). These arguments together with Condition 2 show that all the conditions of Corollary 3.9 of McLeish (1975a) are satisfied. Thus, \( S_T \to_d N(0, \lambda' \Omega_\vartheta \lambda) \). The result now follows from the Cramer-Wold theorem.

Consistency of \( \hat{\theta} \) follows directly from the asymptotic distribution which implies that \( T^{1/2} \left( \hat{\theta} - \theta \right) = O_p(1) \) such that \( \hat{\theta} = \theta + o_p(1) \).

The following theorem establishes the limiting distribution of the test statistic in (5).

**Theorem 2** Assume that Conditions 2, 3, 4, 5 and 6 hold. For \( \nu_t = \nu_t(\xi_0) \) let \( V_t = \nu_t' \nu_t - V \) where \( V \) is a fixed, positive definite matrix. Assume that for any element \( \nu_{t,i} \) of \( \nu_t \), \( E[|\nu_{t,i}|^{p+\varepsilon}] < \infty \) where \( \varepsilon \) is
the same as in Condition 6. Then,

\[ T\hat{m}'\hat{V}^{-1}\hat{m} \rightarrow_d \chi_{(k,k_w)}^2 \]

**Proof.** First consider \( \sum_{t=1}^T \hat{D}_t(z_t, \hat{\psi}) \otimes w_t \) with representative element

\[
\sum_{t=1}^T \delta_{t,j} (\hat{\psi}) w_t = \sum_{t=1}^T \left( \delta_{t,j} (\hat{\psi}) - z_t'(Z'Z)^{-1}Z'\delta_j (\hat{\psi}) \right) w_t
\]

\[
= \sum_{t=1}^T \left( \delta_{t,j} (\hat{\psi}) - \sum_{s=1}^T \delta_{js} (\hat{\psi}) z_s'(Z'Z)^{-1} z_t \right) w_t
\]

\[
= \sum_{t=1}^T \delta_{t,j} (\hat{\psi}) w_t - \sum_{t=1}^T \delta_{js} (\hat{\psi}) z_s' \pi_w
\]

\[
= \sum_{t=1}^T \delta_{t,j} (\hat{\psi}) (w_t - \hat{\pi}_w z_s) .
\]

Thus, the test we consider is based on \( \delta_{t,j} (\hat{\psi}) (w_t - \hat{\pi}_w z_s) \). Recall \( \hat{m}_t = (D_t(z_t, \hat{\psi})) \otimes (w_t - \hat{\pi}_w z_t) \) such that for \( m_t (\xi) = (D_t(z_t, \xi)) \otimes (w_t - \pi_w z_t) \) and \( m_{t,0} = m_t (\xi_0) \) and the mean value theorem it follows that

\[
\hat{m}_t = m_t (\xi_0) + \partial m_t (\xi) / \partial \psi' \left( \hat{\psi} - \psi_0 \right)
\]

with \( \| \xi - \xi_0 \| \leq \| \hat{\xi} - \xi_0 \| . \) Using (3) as well as Condition 4 and setting \( \hat{m}_t = T^{-1} \sum_{t=1}^T \partial m_t (\xi) / \partial \psi' \) we obtain

\[
T^{-1/2} \sum_{t=1}^T \hat{m}_t = T^{-1/2} \sum_{t=1}^T m_{t,0} + \hat{m}_t (\hat{\xi}) \Omega_{\psi}^{-1}T^{-1/2} \sum_{t=1}^T l(D_t, z_t, \psi_0) + (\pi_w - \hat{\pi}_w) T^{-1/2} \sum_{t=1}^T z_t \delta_{t,j} (\hat{\psi}) + o_p (1).
\]

\[
= T^{-1/2} \sum_{t=1}^T \left( m_{t,0} + \hat{m}_t (\xi_0) \Omega_{\psi}^{-1}l(D_t, z_t, \psi_0) \right) + o_p (1)
\]

where the last line follows by the same arguments as in the proof of Theorem 1. With \( \nu_t (\xi_0) = m_t (\xi) + \hat{m}_t (\xi_0) \Omega_{\psi}^{-1}l(D_t, z_t, \psi_0) \) it follows from Corollary 3.9 of McLeish (1975a) that

\[
T^{-1/2} \sum_{t=1}^T \hat{m}_t = T^{-1/2} \sum_{t=1}^T \nu_t (\xi_0) + o_p (1) \rightarrow_d N (0, V)
\]

(17)

where \( V = E \left[ \nu_t (\xi_0) \nu_t (\xi_0)' \right] \) is a \((k \cdot k_w) \times (k \cdot k_w)\) non-singular matrix. A detailed verification of the conditions is omitted but follows the same line of argument as given in the proof of Theorem 1 above. To estimate \( V \), define

\[
\hat{\nu}_t = \hat{m}_t + \hat{m}_t (\hat{\xi}) \Omega_{\psi}^{-1}l(D_t, z_t, \hat{\psi})
\]
with

\[ \hat{\Omega}_\psi = -T^{-1} \sum_{t=1}^{T} \frac{\partial l(D_t, z_t, \hat{\psi})}{\partial \psi'} . \]

Let

\[ \hat{\nu} = T^{-1} \sum_{i=1}^{T} \hat{\nu}_t \hat{\nu}'_t . \]

By arguments similar to the proof of Theorem 1 it follows that

\[ \hat{\Omega}_\psi \to_p \Omega_\psi \] (18)

and

\[ \hat{m} \left( \xi \right) \to_p \hat{m} \left( \xi_0 \right) . \] (19)

Next, expand

\[ \hat{\nu}_t = m_{t,0} + \partial m_t \left( \hat{\xi} \right) / \partial \psi' \left( \hat{\psi} - \psi_0 \right) \]

\[ + \left( \hat{m} \left( \hat{\xi} \right) \hat{\Omega}_\psi^{-1} - \hat{m} \left( \xi_0 \right) \Omega_\psi^{-1} \right) l(D_t, z_t, \hat{\psi}) \]

\[ + \hat{m} \left( \xi_0 \right) \Omega_\psi^{-1} \left( l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right) \]

\[ + \hat{m} \left( \xi_0 \right) \Omega_\psi^{-1} l(D_t, z_t, \psi_0) \]

and recalling \( \nu_t = m_{t,0} + \hat{m} \left( \xi_0 \right) \Omega_\psi^{-1} l(D_t, z_t, \psi_0) \). Then,

\[ \left\| T^{-1} \sum_{t=1}^{T} \hat{\nu}_t \hat{\nu}'_t - V \right\| \leq \left\| T^{-1} \sum_{t=1}^{T} \left( \hat{\nu}_t \hat{\nu}'_t - \nu_t \nu'_t \right) \right\| + \left\| T^{-1} \sum_{t=1}^{T} \nu_t \nu'_t - V \right\| \] (20)

where the second term on the RHS of (20) is \( o_p \left( 1 \right) \) by Theorem 2.10 of McLeish (1995b). Next, consider

\[ T^{-1} \sum_{t=1}^{T} \left( \hat{\nu}_t \hat{\nu}'_t - \nu_t \nu'_t \right) = T^{-1} \sum_{t=1}^{T} \left( \hat{\nu}_t - \nu_t \right) \left( \hat{\nu}_t - \nu_t \right)' + \nu_t \left( \hat{\nu}_t - \nu_t \right)' - \left( \hat{\nu}_t - \nu_t \right) \nu'_t \] (21)

where

\[ \hat{\nu}_t - \nu_t = \partial m_t \left( \hat{\xi} \right) / \partial \psi' \left( \hat{\psi} - \psi_0 \right) + \left( \hat{m} \left( \hat{\xi} \right) \hat{\Omega}_\psi^{-1} - \hat{m} \left( \xi_0 \right) \Omega_\psi^{-1} \right) l(D_t, z_t, \hat{\psi}) \]

\[ + \hat{m} \left( \xi_0 \right) \Omega_\psi^{-1} \left( l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right) . \] (22)
Thus,
\[
T^{-1} \sum_{t=1}^{T} \nu_t (\hat{\nu}_t - \nu_t)' = T^{-1} \sum_{t=1}^{T} \nu_t \left( \frac{\partial m_t (\xi_t)}{\partial \psi'} (\hat{\psi} - \psi_0) \right)' + T^{-1} \sum_{t=1}^{T} \nu_t \left( \left( \hat{m}_t (\xi_t) \hat{\Omega}_\psi^{-1} - \hat{m}_t (\xi_0) \Omega_\psi^{-1} \right) l(D_t, z_t, \hat{\psi}) \right)' + T^{-1} \sum_{t=1}^{T} \nu_t \left( \hat{m}_t (\xi_0) \Omega_\psi^{-1} \left( l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right) \right)'
\]
\[
\equiv R_1 + R_2 + R_3.
\]

For $R_1$ note that
\[
\|R_1\| \leq \left\| T^{-1} \sum_{t=1}^{T} \nu_t \frac{\partial m_t (\xi_0)}{\partial \psi'} \right\| \|\hat{\psi} - \psi_0\| + T^{-1} \sum_{t=1}^{T} \|\nu_t\| \left\| \frac{\partial m_t (\xi_0)}{\partial \psi'} - \frac{\partial m_t (\hat{\xi}_t)}{\partial \psi'} \right\| \|\hat{\psi} - \psi_0\|
\]
\[
\text{where } \|\hat{\psi} - \psi_0\| = O_p \left( T^{-1/2} \right) \text{ and } T^{-1} \sum_{t=1}^{T} \nu_t \frac{\partial m_t (\xi_0)}{\partial \psi'} = O_p (1)
\]
because
\[
E \left[ \|\nu_t \frac{\partial m_t (\xi_0)}{\partial \psi'} \|^{p+\epsilon}/2 \right] \leq \left( E \left[ \|\nu_t\|^{p+\epsilon} \right] E \left[ \left\| \frac{\partial m_t (\xi_0)}{\partial \psi'} \right\|^{p+\epsilon} \right] \right)^{1/2} < \infty
\]
by Condition 6 and by Theorem 2.10 of McLeish (1975b).\footnote{We use McLeish (1975), Equation (2.12) and stationarity to establish Condition (2.11) of Theorem (2.10).} The second term in (24) can be bounded with probability approaching 1 as $T \to \infty$, using Condition 6(iii), and noting that
\[
\left\| \frac{\partial m_t (\xi_0)}{\partial \psi'} - \frac{\partial m_t (\hat{\xi}_t)}{\partial \psi'} \right\| \leq B_t \|\hat{\xi}_t - \xi_0\|^\alpha,
\]
by
\[
T^{-1} \sum_{t=1}^{T} \|\nu_t\| \left\| \frac{\partial m_t (\xi_0)}{\partial \psi'} - \frac{\partial m_t (\hat{\xi}_t)}{\partial \psi'} \right\| \|\hat{\psi} - \psi_0\| \leq \left\| \hat{\xi}_t - \xi_0\right\|^{1+\alpha} T^{-1} \sum_{t=1}^{T} \|\nu_t\| \|B_t\|
\]
\[
\text{where } E \left[ \|\nu_t\|^{p+\epsilon}/2 |B_t|^{(p+\epsilon)/2} \right] \leq \left( E \left[ \|\nu_t\|^{p+\epsilon} \right] E \left[ |B_t|^{p+\epsilon} \right] \right)^{1/2} < \infty \text{ by Condition 6. This again implies that}
\]
\[
T^{-1} \sum_{t=1}^{T} \|\nu_t\| \|B_t\| = O_p (1)
\]
by McLeish (1975b). Now (25) and (26) imply that $R_1 = o_p (1)$.\footnote{We use McLeish (1975), Equation (2.12) and stationarity to establish Condition (2.11) of Theorem (2.10).}
For $R_2$ note that using Condition 6(ii), w.p.a.1 as $T \to \infty$,

$$\|R_2\| \leq \left\| \hat{m} \left( \hat{\xi} \right) \Omega_{\psi}^{-1} - \hat{m} \left( \xi_0 \right) \Omega_{\psi}^{-1} \right\| T^{-1} \sum_{t=1}^{T} \| \nu_t \| \| l (D_t, z_t, \psi_0) \| + \left\| \hat{m} \left( \hat{\xi} \right) \Omega_{\psi}^{-1} - \hat{m} \left( \xi_0 \right) \Omega_{\psi}^{-1} \right\| T^{-1} \sum_{t=1}^{T} \| \nu_t \| \| l (D_t, z_t, \hat{\psi}) \|$$

$$\leq \left\| \hat{m} \left( \hat{\xi} \right) \Omega_{\psi}^{-1} - \hat{m} \left( \xi_0 \right) \Omega_{\psi}^{-1} \right\| T^{-1} \sum_{t=1}^{T} \| \nu_t \| \| l (D_t, z_t, \psi_0) \| + \left\| \hat{m} \left( \hat{\xi} \right) \Omega_{\psi}^{-1} - \hat{m} \left( \xi_0 \right) \Omega_{\psi}^{-1} \right\| T^{-1} \sum_{t=1}^{T} \| \nu_t \| \| \hat{\psi} - \psi_0 \|$$

where $E \left[ (\| \nu_t \| \| l (D_t, z_t, \psi_0) \|)^{(p+\epsilon)/2} \right] < \infty$ as before. Then, $T^{-1} \sum_{t=1}^{T} \| \nu_t \| \| l (D_t, z_t, \psi_0) \| = O_p (1)$ and (18), (19) and (27) imply that $R_2 = o_p (1)$.

For $R_3$ note that

$$\left\| T^{-1} \sum_{t=1}^{T} \nu_t \left( \hat{m} \left( \xi_0 \right) \Omega_{\psi}^{-1} \left( l (D_t, z_t, \hat{\psi}) - l (D_t, z_t, \psi_0) \right) \right) \right\|$$

$$\leq \left\| \hat{m} \left( \xi_0 \right) \Omega_{\psi}^{-1} \right\| T^{-1} \sum_{t=1}^{T} \| \nu_t \| \| l (D_t, z_t, \hat{\psi}) - l (D_t, z_t, \psi_0) \|$$

$$= \left\| \hat{m} \left( \xi_0 \right) \Omega_{\psi}^{-1} \right\| T^{-1} \sum_{t=1}^{T} \| \nu_t \| \| \hat{\psi} - \psi_0 \|$$

where $\| \hat{\psi} - \psi_0 \| = o_p (1)$ by Condition 4. Then, $R_3 = o_p (1)$ follows from (27). The term $T^{-1} \sum_{t=1}^{T} (\hat{\nu}_t - \nu_t) \times (\hat{\nu}_t - \nu_t)'$ in (21) can be analyzed in the same way as $T^{-1} \sum_{t=1}^{T} \nu_t (\hat{\nu}_t - \nu_t)'$ but the details are omitted.

It follows that $T^{-1} \sum_{t=1}^{T} (\hat{\nu}_t \nu'_t - \nu_t \nu'_t) = o_p (1)$ which in turn implies that

$$\hat{V} - V = o_p (1).$$

Then, for $\tilde{m} = T^{-1} \sum_{t=1}^{T} \hat{m}_t$, the statistic $T \tilde{m} \hat{V}^{-1} \tilde{m}$ is asymptotically $\chi^2_{(k-k_w)}$ because of (17), (28) and the continuous mapping theorem. \qed
References


