Asymptotic Inference About Predictive Ability
Kenneth D. West
Last Revised December, 1994

Additional Appendix

This additional appendix contains material omitted from the body of the paper to save space:

I. Additional simulation results:

Plots of Actual versus Nominal Sizes of Hypothesis Tests
   Experiment A, Mean Prediction Error Figure AA1
   Experiment A, Mean Squared Prediction Error Figure AA2
   Experiment A, First Order Serial Correlation Figure AA3
   Experiment B, π=P/R Figure AA4
   Experiment B, π=0 Figure AA5
   Description of Experiment C
   Empirical sizes of nominal .05 tests, experiment C Table AA1

II. Proofs and additional discussion:

A. Lemma A4(a) (additional detail)
B. Lemma A6
C. Lemma 4.3 (additional detail)
D. Theorem 5.1
E. Comment 6 in section 4
F. GLS illustration (footnote 4)
G. Statement of assumptions and results when f depends on \( \hat{\beta}_t, \ldots, \hat{\beta}_{t-L} \) for L>0 (mentioned in section 2)
H. Test for serial correlation in one-step ahead prediction errors (comment 3 in section 4)
I. Test for zero MPE (comment 5 in section 4)

IV. Additional reference
Figure AA2
Actual versus Nominal Sizes of Hypothesis Tests
Experiment A. Mean Squared Prediction Error
Figure AA3
Actual versus Nominal Sizes of Hypothesis Tests
Experiment A. First Order Serial Correlation
Figure AA4: Actual versus Nominal Sizes of Hypothesis Tests
Experiment B. \( \pi = P/R \)
Figure AA5
Actual versus Nominal Sizes of Hypothesis Tests
Experiment B, π=0
Description of Experiment C:

The DGP was $y_t = y_{t-1} + u_t$, $u_t \sim iid \ N(0,1)$, where all variables are scalars. To generate a sample of size 200, $y_0$ was drawn from its unconditional normal distribution and $y_t$ for $t=1,\ldots,200$ was generated recursively. There were 5000 repetitions. Assumption 2(b) presumes that successive samples for regression estimation come from adding an additional observation to the end of the previous sample. I will call this procedure for selecting the parameter estimates the recursive procedure, that underlying (4.6) the fixed procedure, and that underlying (4.7) the rolling procedure.

For given $R$, the series of one step ahead prediction errors for $t=R+1,\ldots,R+P$ was obtained as follows. Let $\hat{\beta}_t$ denote the OLS estimate from a sample going from 1 to $t$. Recursive: $\hat{u}_{t+1} = y_{t+1} - y_t \hat{\beta}_t$. Fixed: $\hat{u}_{t+1} = y_{t+1} - y_t \hat{\beta}_R$.
Rolling: $\hat{u}_{t+1} = y_{t+1} - y_t \hat{\beta}_{t,R}$, where $\hat{\beta}_{t,R}$ is the OLS estimate from a sample going from $t-R+1$ to $t$.

For all three procedures the asymptotic variance-covariance matrix for the one step ahead MSPE is $E\hat{u}_t^4 - (E \hat{u}_t^2)^2$, and was estimated as $(P^{-1} \hat{\Sigma}_{t+1}^4 - (P^{-1} \hat{\Sigma}_{t+1}^2)^2$.

For first order serial correlation:
A. Recursive: $\Omega = \hat{\Omega} = 1$.
B. Fixed: Equation (4.6) simplifies to $\Omega = 1 + \pi \hat{\nu}_R$. I set $\hat{\Omega} = 1 + \frac{\hat{\nu}_R}{R}$, $\hat{\nu}_R = \hat{\sigma}^2 (R^{-1} \Sigma_{t+1}^2)^{-1}$, $\hat{\sigma}^2 = R^{-1} \Sigma_{t+1}^2 (y_{t+1} - y_{t} \hat{\beta}_R)^2$.
C. Rolling: Equation (4.7) simplifies to $\Omega = 1 - \Pi \hat{\nu}_R$, $\Pi = \pi^2/3$ for $\pi \leq 1$, $\Pi = 1 - (2/3\pi)$ for $\pi > 1$. $\hat{\Omega}$ was constructed by setting $\pi = \frac{R}{P}$ and computing $\hat{\nu}_R$ from the last regression estimate (i.e., $\hat{\nu}_R$ was computed as for the fixed procedure except that the sample went from $P$ to $P+R-1$ rather than from 1 to $R$.) The resulting $\hat{\Omega}$ may be negative (just as Durbin's [1970] procedure for in-sample testing for serial correlation may lead to a negative estimated variance). When this happened, the statistic was understood to reject at every significance level.

Panel D reports results for the serial correlation test in the fixed and
rolling procedures when \( \hat{n} \) is set to 1 so that uncertainty about \( \beta^* \) is ignored in testing. By construction, the panel D1 entries are larger than those in panel B2, and the panel D2 entries are smaller than those in panel C2. As expected, the panel D sizes are particularly bad when P/R is relatively large.

### Table AA1

Size of Nominal .05 Tests, Experiment C

<table>
<thead>
<tr>
<th>A. Recursive</th>
<th>2. First Order Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MSPE</td>
<td>P</td>
</tr>
<tr>
<td>R</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>0.105</td>
</tr>
<tr>
<td>50</td>
<td>0.106</td>
</tr>
<tr>
<td>100</td>
<td>0.108</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Fixed</th>
<th>2. First Order Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MSPE</td>
<td>P</td>
</tr>
<tr>
<td>R</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>0.112</td>
</tr>
<tr>
<td>50</td>
<td>0.108</td>
</tr>
<tr>
<td>100</td>
<td>0.109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Rolling</th>
<th>2. First Order Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MSPE</td>
<td>P</td>
</tr>
<tr>
<td>R</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>0.102</td>
</tr>
<tr>
<td>50</td>
<td>0.105</td>
</tr>
<tr>
<td>100</td>
<td>0.109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D. First Order Serial Correlation, ( P\hat{\rho}^2 ) Assumed to be ( \chi^2(1) )</th>
<th>2. Rolling</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fixed</td>
<td>P</td>
</tr>
<tr>
<td>R</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>0.112</td>
</tr>
<tr>
<td>50</td>
<td>0.077</td>
</tr>
<tr>
<td>100</td>
<td>0.054</td>
</tr>
</tbody>
</table>
II. Proofs and additional discussion

A. Lemma A4(a) (additional detail): Let \( v_t \) and \( \gamma_j \) be defined as in the appendix proof of this Lemma. For notational simplicity, assume that the zero mean bivariate process \((v_t, h_t)'\) is fourth order stationary. Let
\[
\kappa(i,j,n) = \text{Ev}_t v_{t-i} h_{t-j} h_{t-n} - \text{Ev}_t v_{t-i} \text{E} h_{t-j} h_{t-n} - \gamma_j \gamma_{n-i} \gamma_n \gamma_j
\]
denote the fourth cumulant \( \kappa \); by assumption 3(a),
\[
\sum_{i,j,n=-\infty}^{\infty} |\kappa(i,j,n)| < \infty
\]
(Andrews (1991)). To show \( \text{var}[P^{-1/2} \Sigma v(t)] \to 0 \): We have

\[
\text{var}[P^{-1/2} \Sigma v(t)] = C_{1T} + C_{2T} + C_{3T},
\]

\[
C_{1T} = P^{-1} \Sigma_{t \in R} P^{-1} \Sigma_{1 \leq t \leq n} R_{t-j}^{-1} [s^{-1} t^{-1} \gamma_{t-i} \gamma_{s-j}],
\]

\[
C_{2T} = P^{-1} \Sigma_{t \in R} P^{-1} \Sigma_{j \in R} R_{t-j}^{-1} [s^{-1} t^{-1} (\text{Ev}_t v_t) (\text{E} h_{t-j})],
\]

\[
C_{3T} = P^{-1} \Sigma_{t \in R} P^{-1} \Sigma_{j \in R} R_{t-j}^{-1} [s^{-1} t^{-1} \kappa(t-s, t-j, t-i)].
\]

Note that \( |C_{1T}| \leq P^{-1} \Sigma_{t \in R} P^{-1} \Sigma_{j \in R} [s^{-1} t^{-1} (\Sigma_{j=s}^{\infty} |\gamma_j|)^2] = (\Sigma_{j=s}^{\infty} |\gamma_j|)^2 P^{-1} (\Sigma_{t \in R} R_{t}^{-1} t^{-1} )^2 \to 0 \) by Lemma A6(a). A similar argument shows \( C_{2T} \to 0 \). In \( C_{3T} \), there are at most \( P \) occurrences of a given triple of values of \( t-s, t-j \) and \( t-i \) \( \implies \)
\[
|C_{3T}| \leq P^{-1} \Sigma_{t \in R} P^{-1} \Sigma_{j \in R} P^{-1} \Sigma_{i \in R} P^{-1} |\kappa(i,j,n)| \to 0.
\]

B. Proof of Lemma A6: When \( \pi=0 \), the result may be established as in the first part of the proof of Lemma A5. I will sketch the argument when \( \pi \neq 0 \). For notational simplicity, in this proof only, assume \( q=1 \), let
\[
u_{t+1} = (f_{t+1} - \text{Ef}_t),
\]
\[
\gamma_j = \text{Ev}_t h_{t-j} \implies S_{hn} = \Sigma_{j=s}^{\infty} \gamma_j.
\]
The aim, then, is to show that
\[
P^{-1} E(u_{R+1} + \ldots + u_{R+p}) \Sigma H(t) = B_1 + B_2 \to [1 - \pi^{-1} \ln(1+\pi)] \Sigma_{j=s}^{\infty} \gamma_j,
\]

\[
B_1 = P^{-1} a_{R,0} E(u_{R+1} + \ldots + u_{R+p}) (h_{1} + \ldots + h_{R}),
\]

\[
B_2 = P^{-1} E(u_{R+1} + \ldots + u_{R+p}) (a_{R,1} h_{R+1} + \ldots + a_{R,p} h_{R+p-1}),
\]
where \(a_{R,j}\) is defined as in the proof of Lemma A5. I will show that \(B_1 \to 0\), \(B_2 \to [1-\pi^{-1}\ln(1+\pi)]\Sigma_{j=0}^{\infty}a_{j+1}\).

\(B_1\): We have

\[(B.1) \quad E(u_{R+1} + \ldots + u_{R+p})(h_1 + \ldots + h_R) =
\gamma_1 + \gamma_2 + \ldots + \gamma_R
+ \gamma_2 + \ldots + \gamma_R + \gamma_{R+1}
+ \ldots + \gamma_p + \gamma_{p+1} + \ldots + \gamma_{R+p-1} =
\gamma_1 + 2\gamma_2 + \ldots + P(\gamma_p + \gamma_{p+1} + \ldots + \gamma_R) + (P-1)\gamma_{R+1} + \ldots + \gamma_{R+p-1}\] if \(P \leq R\)
\[= \gamma_1 + 2\gamma_2 + \ldots + R(\gamma_p + \gamma_{p+1} + \ldots + \gamma_p) + (R-1)\gamma_{R+1} + \ldots + \gamma_{R+p-1}\] if \(P > R\)

\[\implies |B_1| \leq p^{-1}a_{R,0}\Sigma_{j=0}^{\infty}|\gamma_j| \to 0 \text{ by Lemma A2(b)}.\]

\(B_2\): We have

\[E(u_{R+1} + \ldots + u_{R+p})(a_{R,1}h_{R+1} + \ldots + a_{R,p-1}h_{R+p-1}) =
\Sigma_{j=0}^{p-1}(c_j - c_0)\gamma_j + \Sigma_{j=0}^{p-1}(c_j - c_0)\gamma_j,
\] where \(c_0 = a_{R,1} + \ldots + a_{R,p-1},\)
\(c_1 = a_{R,1} + \ldots + a_{R,p-1}, c_2 = a_{R,1} + \ldots + a_{R,p-2}, \ldots, c_{p-1} = a_{R,1},\)
\(c_{-1} = a_{R,2} + \ldots + a_{R,p-1}, c_{-2} = a_{R,3} + \ldots + a_{R,p-1}, \ldots, c_{-p+2} = a_{R,p-1}.\)

By logic such as that in the proof of Lemma A5,

\[P^{-1}c_0 \sim P^{-1}\int_1^{p-1}[\int_T^{p-1}(R+x)^{-1}dx]dy \to [1-\pi^{-1}\ln(1+\pi)].\]

where "\(\sim\)" means "is arbitrarily close to for sufficiently large \(T.\)" The desired result follows if

\[P^{-1}\Sigma_{j=1}^{p-1}(c_j - c_0)\gamma_j \to 0, \quad P^{-1}\Sigma_{j=0}^{p-2}(c_j - c_0)\gamma_j \to 0.\]

Since \(c_j \leq c_0\) and \(P^{-1}c_0\) is bounded, \(|P^{-1}\Sigma_{j=1}^{p-1}(c_j - c_0)\gamma_j| \leq 2P^{-1}c_0|\Sigma_{j=1}^{p-1}\gamma_j| \to 0.\)

For \(j \leq R,\) use

\[|c_j - c_0| \leq R^{-1}[1+2+\ldots(j-1)] = .5R^{-1}(j-1) \leq .5j \quad \text{for } j > 0\]
\[ |c_j - c_0| \leq |j| (P/R) \quad \text{for } j < 0 \]

\[ \implies |P^{-1}\Sigma_{j \in \mathbb{R}}(c_j - c_0)\gamma_j| \leq 0.5P^{-1}\Sigma_{j=1}^{\infty}|j|\gamma_j + R^{-1}\Sigma_{j=1}^{\infty}|j|\gamma_j \rightarrow 0 \]

by Lemma A2(a).

C. Additional detail on proof of Lemma 4.3: To show $V(\pi)$ is positive definite:

define the following $(l+q) \times (l+q)$ matrices:

\[
S = \begin{pmatrix} S_{ef} & S_{ef} B' \\ (BS_{ef}' \quad V_{\beta} ) \end{pmatrix}, \quad A_1 = \begin{pmatrix} (I_t \quad 0) \\ (0 \quad \Pi_{k}) \end{pmatrix}, \quad A_2 = \begin{pmatrix} (0 \quad 0) \\ (0 \quad (2\Pi - \Pi^2)V_{\beta}) \end{pmatrix}. 
\]

Then $V(\pi) = A_1 S A_1' + A_2$. By assumption, $S$ is positive definite; since $0 \leq \Pi \leq 1$, $\Pi^2 < 2\Pi$ so $A_2$ is positive semidefinite.

D. Proof of Theorem 5.1: (a) Consider $\hat{\Gamma}_{ef}(0) = P^{-1}\Sigma[f_{t+r}(\hat{\beta}_t) - \bar{f}]^2$; other autocovariances may be handled similarly. Expand $f_{t+r}(\hat{\beta}_t)$ around $\hat{\beta}^*$_t:

\[ f_{t+r}(\hat{\beta}_t) - \bar{f} = \phi_{t+r} + r_{t+r}, \]

(D.1) \hspace{1cm} \phi_{t+r} = f_{t+r}(\hat{\beta}^* - Ef_t), \quad r_{t+r} = (Ef_t - \bar{f}) + [f_{t+r, \hat{\beta}(\hat{\beta}^* - \bar{\beta})} + w_{t+r}].

\[ \implies \]

(D.2) \hspace{1cm} \hat{\Gamma}_{ef}(0) = P^{-1}\Sigma \phi^2_{t+r} + 2P^{-1}\Sigma[\phi_{t+r} r_{t+r}] + P^{-1}\Sigma r^2_{t+r}

(Recall that I am assuming that $l=1$ for notational simplicity.) The first term on the right hand side of (D.2) converges in probability to $\Gamma_{ef}(0)$ by White (1984, Corollary 3.48). The second term on the r.h.s. of (D.2) is 2 times

(D.3) \hspace{1cm} (Ef_t - \bar{f})P^{-1}\Sigma \phi_{t+r} + P^{-1}\Sigma[\phi_{t+r, \hat{\beta}(\hat{\beta}^* - \bar{\beta})}] + P^{-1}\Sigma(\phi_{t+r} w_{t+r})

The first term on the r.h.s. of (D.3) is $o_p(1)$ by $\bar{f} \overset{P}{\rightarrow} Ef_t$ and White (1984, Corollary 3.48). The absolute value of the second term on the r.h.s. of (D.3) is bounded above by
\[ \sup_t |\hat{\beta}_t - \beta^*| \left( P^{-1} \Sigma \hat{f}_{t+\tau} \right)^{1/2} \left( P^{-1} \Sigma |f_{t+\tau, \beta}|^2 \right)^{1/2}, \]

which is \( o_p(1) \) because the first term is \( o_p(1) \) by Lemma A3(b), the second and third are \( O_p(1) \) by Markov's inequality. Since \( w_{t+\tau} = .5(\hat{\beta}_t - \beta^*)^2[\delta^2 f_{t+\tau}(\hat{\beta}_t)/\delta \beta^2] \) for some \( \tilde{\beta}_t \) between \( \hat{\beta}_t \) and \( \beta^* \), the absolute value of the third term on the r.h.s. of (D.3) is bounded above by

\[ \sup_t |(\hat{\beta}_t - \beta^*)|^2 \left( P^{-1} \Sigma \hat{f}_{t+\tau} \right)^{1/2}\left( P^{-1} \Sigma |\delta^2 f_{t+\tau}(\hat{\beta}_t)/\delta \beta^2|^2 \right)^{1/2}, \]

the first term of which is \( o_p(1) \) by Lemma A3(b), the next two of which are \( O_p(1) \) by Assumption 1, Assumption 3 and Markov's inequality. Finally, the third term on the r.h.s. of (D.2) is

\[
P^{-1}(Ef_t - \tilde{f})^2 + 2(Ef_t - \tilde{f})P^{-1} \Sigma[f_{t+\tau, \beta}(\hat{\beta}_t - \beta^*)] + 2(Ef_t - \tilde{f})P^{-1} \Sigma w_{t+\tau}\]

\[ + (P^{-1} \Sigma[f_{t+\tau, \beta}(\hat{\beta}_t - \beta^*)])^2 + 2P^{-1} \Sigma[f_{t+\tau, \beta}(\hat{\beta}_t - \beta^*)]w_{t+\tau} + P^{-1} \Sigma w_{t+\tau}^2. \]

That each of the six terms converges in probability to zero follows from \( \tilde{f} \overset{p}{\rightarrow} Ef_t \), Lemma A3(b), and Assumptions 1 and 3, using logic similar to that just applied to the third term on the r.h.s. of (D.3).

**Lemma 1(b)** Note that even when \( \pi = \infty \), \( M/P^a \rightarrow 0 \) since \( M/R^a \rightarrow 0 \) \( \implies (P/R)^a(M/P^a) \rightarrow 0 \).

The result follows if \( \hat{S}_{hh} \overset{P}{\rightarrow} S_{hh} \), \( \hat{S}_{ff} \overset{P}{\rightarrow} S_{ff} \) and \( \hat{S}_{th} \overset{P}{\rightarrow} S_{th} \). That \( \hat{S}_{hh} \overset{P}{\rightarrow} S_{hh} \) follows from Andrews (1991). I will sketch the proof that \( \hat{S}_{ff} \overset{P}{\rightarrow} S_{ff} \); that \( \hat{S}_{th} \overset{P}{\rightarrow} S_{th} \) follows from similar logic.

Let \( k_j = k(j/M) \), suppressing for simplicity dependence of \( k_j \) on \( M \) and thus \( P \). I will show that \( \Sigma_{j=0}^{P-1} k_j \hat{\Gamma}_{ff}(j) \overset{P}{\rightarrow} \Sigma_{j=0}^{\infty} \Gamma_{ff}(j) \), from which it will also follow that \( \Sigma_{j=0}^{P-1} k_j \hat{\Gamma}_{ff}(-j) \overset{P}{\rightarrow} \Sigma_{j=0}^{\infty} \Gamma_{ff}(-j) \) and hence \( \hat{S}_{ff} \overset{P}{\rightarrow} S_{ff} \). Notation: in this proof only,

\[ \Sigma_j = \Sigma_{j=0}^{P-1}. \]

Define \( \phi_{t+\tau} \) and \( \tau_{t+\tau} \) as in (D.1). An expansion such as in (D.2) then yields

\[
\Sigma_j k_j \hat{\Gamma}_{ff}(j) = \Sigma_j k_j (P^{-1} \Sigma_{t=R+j} [\phi_{t+\tau} \phi_{t+\tau-r-j}]) + \Sigma_j k_j (P^{-1} \Sigma_{t=R+j} [\phi_{t+\tau} \tau_{t+\tau-r-j}])
\]

\[ + \Sigma_j k_j (P^{-1} \Sigma_{t=R+j} [\phi_{t+\tau-r-j} \tau_{t+\tau}]) + \Sigma_j k_j (P^{-1} \Sigma_{t=R+j} [\tau_{t+\tau} \tau_{t+\tau-r-j}]) \]
It follows from Andrews (1991) that $\Sigma_j k_j (P^{-1}\Sigma_{t=Rj}[\phi_{t+r}\phi_{t+r-j}]) \overset{P}{\rightarrow} \Sigma_j=0 \Gamma_{fe}(j)$, so it suffices to show that the other three double summations converge in probability to zero. I will show this for the second double summation; the argument for the third and fourth double summations are similar.

By the definition of $r_t$,

\begin{equation}
(D.4) \quad |\Sigma_j k_j (P^{-1}\Sigma_{t=Rj}[\phi_{t+r}r_{t+r-j}])| = |(Ef_t \cdot f)\Sigma_j k_j (P^{-1}\Sigma_{t=Rj}[\phi_{t+r}r_{t+r-j}]) + \Sigma_j k_j (P^{-1}\Sigma_{t=Rj}[\phi_{t+r}w_{t+r-j}])| \\
\end{equation}

We have

$$
|P^{-1}\Sigma_{t=Rj}[\phi_{t+r}f_{t+r-j},\beta]| \leq (P^{-1}\Sigma_{t=Rj}\phi_{t+r}^2)^{1/2}(P^{-1}\Sigma_{t=Rj}|f_{t+r-j},\beta|^2)^{1/2} \\
\leq (P^{-1}\Sigma_{t=Rj}\phi_{t+r}^2)^{1/2}(P^{-1}\Sigma_{t=Rj}|f_{t+r},\beta|^2)^{1/2},
$$

and, similarly, for $m_t$ defined in the statement of Theorem 5.1,

$$
|P^{-1}\Sigma_{t=Rj}[\phi_{t+r}w_{t+r-j}]| = .5 |P^{-1}\Sigma_{t=Rj}(\phi_{t+r}(\hat{\beta}_{t-j}-\beta^*)^2[\beta^2f_{t+r-j}(\hat{\beta}_{t-j})/\beta^2])| \leq .5 \sup_t |\hat{\beta}_{t-j}-\beta^*|^2(P^{-1}\Sigma_{t=Rj}\phi_{t+r}^2)^{1/2}(P^{-1}\Sigma_{t=Rj}m_{t+r}^2)^{1/2},
$$

Let $\alpha$ be defined as in the statement of the Theorem. Then for $\pi<\infty$ (D.4) is bounded above by

\begin{equation}
(D.5) \quad (M/P^\alpha)(M^{-1}\Sigma_j|k_j|) + |P^\alpha(Ef_t \cdot f)|P^{-1}\Sigma_{t=Rj}|\phi_{t+r}| + \\
P^\alpha\sup_t |\hat{\beta}_{t-j}-\beta^*|(P^{-1}\Sigma_{t=Rj}\phi_{t+r}^2)^{1/2}(P^{-1}\Sigma_{t=R}|f_{t+r},\beta|^2)^{1/2} + \\
.5 P^\alpha\sup_t |\hat{\beta}_{t-j}-\beta^*|^2(P^{-1}\Sigma_{t=Rj}\phi_{t+r}^2)^{1/2}(P^{-1}\Sigma_{t=Rj}m_{t+r}^2)^{1/2}.
\end{equation}

By Theorem 4.1 $|P^\alpha(Ef_t \cdot f)| \overset{P}{\rightarrow} 0$. Since $\pi<\infty$, it follows from Lemma A3(b) that $P^\alpha\sup_t |\hat{\beta}_{t-j}-\beta^*| \overset{P}{\rightarrow} 0$. By Markov's inequality each of the summations inside the braces is $O_p(1)$. By assumption, $M/P^\alpha \rightarrow 0$ and $M^{-1}\Sigma_j|k_j| + \int_0^\infty |k(x)|dx < \infty$, so the desired result follows.

For $\pi=\infty$, the logic is the same except that in (D.5) $P^\alpha$ is replaced by $R^\alpha$. 
E. Comment 6 in Section 4

(a) Equation (4.6): Using logic similar to that in the paper, a simple mean value expansion yields

\[ P^{1/2}(\bar{f} - E\bar{f}) = P^{1/2}(f_{t+\tau} - E\bar{f}) + P^{1/2}(P^{-1/2}\Sigma f_{t+\tau,\beta})B(R)H(R) + o_p(1) \]

\[ = P^{-1/2}\Sigma(f_{t+\tau} - E\bar{f}) + (P^{-1}\Sigma f_{t+\tau,\beta})B(R)(P_R^R)[P^{-1/2}\Sigma h_s] + o_p(1) \]

Logic such as that used in the proof of Lemma A6 may be used to show

\[ E P^{-1}\Sigma(f_{t+\tau} - E\bar{f})\Sigma_h^R s' \rightarrow 0, \text{ from which it follows that} \]

\[ [P^{-1/2}\Sigma(f_{t+\tau} - E\bar{f})', (P_R^R)^{1/2} R^{-1/2}\Sigma h_s] \] is asymptotically normal with a block diagonal covariance matrix of \( \text{diag}(S_{ff}, \pi S_{hh}) \). Since \( P^{-1}\Sigma f_{t+\tau,\beta} \overset{p}{\rightarrow} F, \ B(R) \overset{p}{\rightarrow} B \), (4.6) follows.

(b) Equation (4.7): Write the estimate as \( \hat{\beta}_{t,\tau} = B_R(t)H_R(t) \), where e.g. for OLS applied to \( y_t = X_t'\beta^* + u_t \) we have \( B_R(t) = (R^{-1}\Sigma_{s=t}^{t+1}X_sX_s')^{-1}, \ H_R(t) = R^{-1}\Sigma_{s=t}^{t+1}X_su_s \).

The result requires that

\[ \sup_{s \geq t, \tau > 0} |B_R(t) - B| \overset{P}{\rightarrow} 0. \]

When \( B_R(t) \) is a moment matrix such as in the OLS example this may be shown to follow from mixing conditions; see the discussion of E.1 below.

By an argument such as that in the paper, the result will follow from

\[ (E.1) \quad P^{1/2}(\bar{f} - E\bar{f}) = P^{1/2}(f_{t+\tau} - E\bar{f}) + P^{1/2}FB\Sigma H_R(t) + o_p(1). \]

\[ (E.2) \quad \lim E[P^{-1}\Sigma H_R(t)\Sigma H_R(t)'] = (\pi - \frac{\pi^2}{3})S_{hh} \quad \pi \leq 1, \]

\[ = (1 - \frac{1}{2\pi})S_{hh} \quad \pi > 1. \]

\[ (E.3) \quad \lim E[P^{-1}\Sigma(f_{t+\tau} - E\bar{f})\Sigma H_R(t)'] = \frac{\pi}{2}S_{fh} \quad \pi \leq 1, \]

\[ = (1 - \frac{1}{2\pi})S_{fh} \quad \pi > 1. \]

I will sketch parts of the proofs of (E.1) and (E.2). The proof of (E.3) and the remaining steps in the proofs of (E.1) and (E.2) are conceptually similar to proofs that appear in the paper.
The proof of (E.1) uses an argument quite similar to that used to establish Lemmas 4.1 and 4.2, except that Lemma A3 is replaced with: For $0 \leq a < 0.5$:

(a) $\sup_{t} |P^a \hat{H}_R(t)| \overset{P}{\to} 0$, (b) $\sup_{t} |P^a(\hat{\beta}_{t,R} - \beta^*)| \overset{P}{\to} 0$. To show (a) (given (a), the proof of (b) is similar to that of Lemma A3(b)): We have

$$H_R(t) = R^{-1}(h_{t-R+1} + \ldots + h_t) = \left(\frac{t}{R}\right)[t^{-1}(h_1 + \ldots + h_t)] - \left(\frac{P}{R}\right)[P^{-1}(h_1 + \ldots + h_{t-R})]$$

$$\implies \sup_{t \leq R+P-1} |P^a H_R(t)|$$

$$\leq \sup_{t \leq R+P-1} \left|\left(\frac{t}{R}\right) P^a H(t)\right| + \sup_{t \leq R+P-1} \left|\left(\frac{P}{R}\right) P^a P^{-1}(h_1 + \ldots + h_{t-R})\right|$$

$$= \sup_{t \leq R+P-1} \left|\left(\frac{t}{R}\right) P^a H(t)\right| + \sup_{t \leq P-1} \left|\left(\frac{P}{R}\right) P^{a-1}(h_1 + \ldots + h_s)\right|$$

Since $\pi < \infty$, for some $b < \infty$ we have $\frac{t}{R} < b < \infty$ for $t \leq R+P-1$ $\implies$ it suffices to show $\sup_{t} |P^a H(t)| \overset{P}{\to} 0$, $\sup_{t \leq P-1} |P^{a-1}(h_1 + \ldots + h_s)| \overset{P}{\to} 0$. The first follows from Lemma A3(a). The second: in light of the mixing property established in the proof of Lemma A3(a), we have $E \left[\sup_{t \leq P-1} (h_1 + \ldots + h_s)^2\right] \leq cP$ for a constant $c$ (Hall and Heyde (1980, p20)) $\implies$ $\sup_{t \leq P-1} |P^{a-2}(h_1 + \ldots + h_s)^2| \overset{P}{\to} 0$ by Markov's inequality $\implies$ $\sup_{t \leq P-1} |P^{a-1}(h_1 + \ldots + h_s)| \overset{P}{\to} 0$.

To show (E.2): Consider the case where $P \leq R$ for all $T$ sufficiently large. The case where $P > R$ is similar, and the case where $P$ fluctuates above and below $R$ indefinitely ($\implies \pi = 1$) follows from the $P \leq R$ and $P > R$ case. For $P \leq R$ we have

$$S_R(t) = R^{-1}(h_1 + 2h_2 + \ldots + (P-1)h_{P-1})$$

$$+ PR^{-1}(h_2 + \ldots + h_R)$$

$$+ R^{-1}[(P-1)h_{R+1} + \ldots + 2h_{R+P-2} + h_{R+P-1}]$$

$$= A_1 + A_2 + A_3.$$

It is easy to see that $\lim \text{var}(P^{-1/2} A_2) = \lim P(R-P+1)R^{-2}E|h_1 + \ldots + P\gamma_1| + o(1) + (\pi - \pi^2)S_{hh}$. I will sketch the argument that shows $\lim \text{var}(P^{-1/2} A_1) = (\pi^2/3)S_{hh}$. That $\lim \text{var}(P^{-1/2} A_3) = (\pi^2/3)S_{hh}$ follows from a nearly identical argument. Since,
finally, it can be shown that \( \lim \text{cov}(P^{-1/2}A_i, P^{-1/2}A_j) = 0 \) for \( i \neq j \), the result will
follow.

For simplicity, assume \( q = 1 \). Let \( \gamma_j = \text{E} h_t h_{t-j} \). Then

\[
\text{var}(A_i) = R^{-2} \sum_{j=1}^{p-2} d_j \gamma_j = R^{-2} d_0 \Sigma \gamma_j - R^{-2} \Sigma (d_0 - d_j) \gamma_j,
\]

\[
d_0 = 1 + 2^2 + \ldots + (P-1)^2,
\]

\[
d_1 = 1.2 + 2.3 + \ldots + (P-2)(P-1),
\]

\[
\ldots
\]

\[
d_{P-2} = 1. (P-1),
\]

\[
d_j = d_j \text{ for } j < 0.
\]

We have \( P^{-1} R^{-2} d_0 \sim \lfloor P^3/(3P^2) \rfloor \to \pi^2/3 \), and the result will follow if \( P^{-1} R^{-2} \Sigma (d_0 - d_j) \gamma_j \to 0 \). This result may be established using \( d_0 \leq \int_0^\infty x^2 dx \), \( d_j \geq \int_{j+1}^{P-1} (x-j) x dx \) \( \to \)

\[
|d_0 - d_j| \leq \left| \int_0^\infty x^2 dx - \int_{j+1}^{P-1} (x-j) x dx \right|, \text{ solving the integrals and manipulating the result to obtain } P^{-1} R^{-2} |\Sigma (d_0 - d_j) \gamma_j| \leq (1/3P) \Sigma |j||\gamma_j| + o(1) \to 0.
\]

F. GLS illustration (footnote 4)

It is evident how to bring certain GLS estimators into the framework of the paper. An example is seemingly unrelated regressions: SUR is simply 3SLS, and so is illustrated in line 2 of Table II. More generally, as noted in footnote 4, GLS estimators fit into the framework of the paper by interpreting them as sequential method of moments estimators. To illustrate this, consider using Durbin's method to make an AR(1) correction. Let

\[
y_t = x_t \delta + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \quad |\rho|<1,
\]

where all variables are scalars. Overly strong conditions to insure that the paper's assumptions are met are that \( \epsilon_t \sim \text{iid normal} \), \( \text{E} \epsilon_t x_s = 0 \) all \( t \) and \( s \) and that \( x_t \sim \text{normal} \).
In a sample of size t, let \( \hat{\rho}_t \) be estimated from the OLS regression

\[
y_s = \rho y_{s-1} + \gamma_1 x_s + \gamma_2 + \epsilon_t = Z_s' \gamma + \epsilon_t,
\]

\[
(\hat{\rho}_t, \hat{\gamma}_{1t}, \hat{\gamma}_{2t})' = (\Sigma Z_s Z_s')^{-1} \Sigma Z_s y_s.
\]

Let

\[
x_s = x_s - \hat{\rho}_t x_{s-1}, \quad y_s = y_s - \hat{\rho}_t y_{s-1},
\]

where the fact that \( \hat{x}_s \) and \( \hat{y}_s \) depend on sample size \( t \) is suppressed for simplicity. The GLS estimate of \( \delta \) is then

\[
\hat{\delta}_t = (\Sigma \hat{x}_s^2)^{-1} \Sigma \hat{x}_s \hat{y}_s.
\]

where the sums run from 2 to \( t \) rather than from 1 to \( t \) but to keep the algebra relatively uncluttered I ignore that asymptotically unimportant fact.

Write the \((4 \times 1)\) population orthogonality condition \( h_t \) used to identify the \((2 \times 1)\) parameter vector \( \beta^* = (\delta, \rho)' \) as

\[
h_t = [(x_t - \rho x_{t-1}) \epsilon_t, Z_t' \epsilon_t]'
\]

with sample counterpart

\[
h_s(\hat{\theta}_t) = [(x_s - \hat{\rho}_t x_{s-1}) [y_s - \hat{\rho}_t y_{s-1} - (x_s - \hat{\rho}_t x_{s-1}) \hat{\delta}_t], Z_s' [y_s - \hat{\rho}_t y_{s-1} - \hat{\gamma}_{1t} x_s - \hat{\gamma}_{2t} x_{s-1}]']
\]

where \( \theta = (\delta, \rho, \gamma_1, \gamma_2)' = (\beta^*, \gamma_1, \gamma_2)' \). A mean value expansion of \( t^{-1} \Sigma h_s(\theta) \) around \( \hat{\theta}_t \) gives

\[
(F.1) \quad H(t) = t^{-1} \Sigma h_s(\theta) = t^{-1} \Sigma h_s(\hat{\theta}_t) + C(\hat{\theta}_t)(\theta - \hat{\theta}_t) = C(\hat{\theta}_t)(\theta - \hat{\theta}_t).
\]

where \( C(\hat{\theta}_t) \) is the \( 4 \times 4 \) matrix of partial derivatives and \( \hat{\theta}_t = (\hat{\delta}_t, \hat{\rho}_t, \hat{\gamma}_{1t}, \hat{\gamma}_{2t})' \) lies between \( \hat{\theta}_t \) and \( \theta \). We then get \( \hat{\theta}_t - \theta = -C(\hat{\theta}_t)^{-1} H(t) \). Write the first two of
this set of four equations as

\[ \hat{\beta}_t - \beta^* = B(t)H(t) \]

where \( \hat{\beta}_t = (\hat{\delta}_t, \hat{\rho}_t)' \) and \( B(t) \) is 2x4. Upon explicitly computing \( C(\hat{\theta}_t) \) and using the partitioned inverse formula we get

\[
B_{11}(t) = [\Sigma(t-1)X_e - \hat{\rho}_tX_{e-1}]^{-1} \\
B_{12}(t) B_{13}(t) B_{14}(t) = B_{11}(t)[-C_{12}(t) 0 0] (\Sigma(t-1)X_e - \hat{\rho}_tX_{e-1})^{-1} \\
\]

\[
C_{12}(t) = -t^{-1}[\Sigma(t-1)X_e - \hat{\rho}_tX_{e-1}]X_{e-1}' - t^{-1}[\Sigma(t-1)X_e - \hat{\rho}_tX_{e-1}] \delta_t X_{e-1} \\
B_{21}(t) = 0 \\
B_{22}(t) B_{23}(t) B_{24}(t) = (1 0 0) (\Sigma(t-1)X_e - \hat{\rho}_tX_{e-1})^{-1} \\
\]

It follows that \( B(t) \overset{a.s.}{=} B \), where \( B \) is the rank 2 matrix

\[
\begin{pmatrix}
(E(x_e - \rho x_{e-1})^2 & 0 & 0 & 0 \\
(0 & (1 0 0)(EZ_e X_e')^{-1})
\end{pmatrix}
\]

(Recall that in this example I am making the textbook assumption that \( Ex_e \epsilon_e = 0 \) for all \( t \) and \( s \).) It is easily verified that \( V_{\beta} = B \Sigma (h_0 \beta') B' \) has rank 2 as well.

G. A. Statement of assumptions and results when \( \bar{f} \) depends on \( \hat{\beta}_t, \ldots, \hat{\beta}_{t-L} \) for \( L \geq 0 \) (mentioned in section 2)

For some fixed and finite integer \( L \geq 0 \), let the estimate of \( Ef_\theta \) depend on \( \hat{\beta}_t, \hat{\beta}_{t-1}, \ldots, \hat{\beta}_{t-L} \):

\[ \bar{f} = \Sigma_{t=0}^{t=L} f_{t+t} (\hat{b}_t), \quad \hat{b}_t \in R^{(L+1)k} = (\beta_1', \beta_{t-1}', \ldots, \beta_{t-L}')', \quad f_{t} : R^{(L+1)k} \rightarrow R^f. \]

When \( L=0 \), as is assumed in the paper, \( \hat{b}_t = \hat{\beta}_t \). Also define the following:

\[ b^* = (\beta^*, \ldots, \beta^*)' \in R^{(L+1)k}, \]

\[ f_{tb} = \frac{\partial f_{t}(b^*)}{\partial b} = [f_{tb0} f_{tb1} \ldots f_{tbl}], \quad l \times k \quad l \times k \quad l \times k \]
\[ f_t(\beta) = f_t[(\beta', \beta', \ldots, \beta')'] \rightarrow f_t : \mathbb{R}^k \rightarrow \mathbb{R}^l; \quad f_{tb} = \sum_{j=0}^{L} f_{tbj}, \quad F = E f_{tb}; \]

\( f_{tbj} \) is the \( l \times k \) matrix that occupies columns \( jk+1 \) to \((j+1)k\) of the \( [l \times (L+1)k] \) matrix \( \partial f_t(b^*)/\partial b \). As the notation suggests, \( f_{tb} = \partial f_t(\beta^*)/\partial \beta \). When \( L=0 \), \( f_{tb} = f_{tb} \) and \( f_t(\beta) = f_t(\beta^*) \) for \( \beta \in \mathbb{R}^k \).

Theorem 4.1 follows when assumption 1 is modified as follows:

**Assumption 1:** In some open neighborhood \( N \) around \( \beta^* \), and with probability one:

(a) For \( b \in \mathbb{N}^{L+1} = [N \times N \times \ldots \times N] \), \( f_t(b) \) is twice continuously differentiable, admitting a mean value expansion

\[ f_t(b) = f_t(b^*) + [\partial f_t(b^*)/\partial b](b-b^*) + \omega_t, \]

where \( \partial f_t(b^*)/\partial b \) is \( l \times (L+1)k \), the \( i' \)th element of the \( (l \times 1) \) vector \( \omega_t \) is

\[ .5(b-b^*)' (\partial^2 f_t(b^*)/\partial b \partial b')(b-b^*) , \]

the scalar \( f_{it}(b) \) is the \( i' \)th component of \( f_t(b) \), and \( \tilde{b} \) is on the line between \( b \) and \( b^* \).

(b) There is a constant \( D < \infty \) such that for all \( t \): (i) \( E |\partial f_t(\beta^*)/\partial b|^2 < D \), and

(ii) for \( i=1, \ldots, l \), \( \sup_{b \in \mathbb{N}^{L+1}} |\partial^2 f_t(b)/\partial b \partial b'| < m_t \) for some \( m_t \) for which \( E m_t < D \).

The generalization of Lemma 5.1 is:

**Lemma 5.1** Let \( \hat{f}_{t, \tau, b} \) denote the \( l \times k \) matrix that occupies columns \( jk+1 \) to \((j+1)k\) of the \( [l \times (L+1)k] \) matrix \( \partial f_{t+\tau}(\hat{b}_t)/\partial b \), \( \partial f_{t+\tau}(\hat{b}_t)/\partial b = [\hat{f}_{t, \tau, b_0} \hat{f}_{t, \tau, b_1} \ldots \hat{f}_{t, \tau, b_L}] \), and define \( \hat{F} = \sum_{j=0}^{L} [P^{-1} \Sigma_{t=0}^{T} \partial f_{t+\tau}(\hat{\beta}_t)/\partial \beta \text{ when } L=0] \). Then \( \hat{F} \rightarrow F \).

**H. Test for serial correlation in one-step ahead prediction errors (comment 3 in section 4)**

A formal statement of the result is:

Let a linear simultaneous equations model be given as \( C y_t = D Z_t + u_t \), where \( y_t \) is the \( l \times 1 \) vector of endogenous variables, \( Z_t \) the \( (r \times 1) \) vector of predetermined variables, and exactly \( k \) elements of \( C \) and \( D \) are not known a priori. Let \( S \) be
p.d., let $C$ be of full rank, and let $v_t = y_t - C^{-1}DZ_t = C^{-1}u_t$. Consider for concreteness the first element of $v_t$, and let $f_t = v_{1t}v_{1t-1}$. Suppose that $E(u_t|Z_t, u_{t-1}, Z_{t-1}, \ldots) = 0$, $E(u_tu_t'|Z_t, u_{t-1}, Z_{t-1}, \ldots) = Eu_tu_t'$. Let $\beta_t$ be estimated by 3SLS (line (2) in Table II). Let $\hat{C}_t$ and $\hat{D}_t$ be estimates constructed from $\beta_t$, $\hat{v}_{t+1} = y_{t+1} - C_t\hat{D}_tZ_{t+1}$, $\hat{v}_{1t+1}$ the first element of $\hat{v}_{t+1}$, $\gamma_j = P^{-1}\Sigma\hat{v}_{1t+1}\hat{v}_{1t+1-j}$, $\hat{\rho} = \gamma_1/\gamma_0$. Then $P^{1/2}\hat{\rho} A N(0,1)$.

I. Test for zero MPE (comment 5 in section 4)

A formal statement of the result is:

Under technical conditions given below, let $y_t = g(X_t, \beta^*) + u_t$, where $y_t$ and $u_t$ are $l \times 1$, and let $\hat{\beta}_t$ be estimated by GMM with a vector of instruments $Z_t$ as described in line (3) of Table II.

(a) Let $y_{t+r}$ be predicted as $g(X_{t+r}, \hat{\beta}_t)$, $\tilde{f} = P^{-1}\Sigma[y_{t+r} - g(X_{t+r}, \hat{\beta}_t)]$. Then $P^{1/2}\tilde{f} A N(0, S_{ff})$, $S_{ff} = \Sigma_{j=0}^{\infty}Eu_{t-j}'$.

(b) Suppose that the model is linear, $y_t = X_t'\beta^* + u_t$, and may be rewritten as $Cy_t = DZ_t + u_t$, where $C$ is of full rank and exactly $k$ elements of $C$ and $D$ are not known a priori. Let $\hat{C}_t$ and $\hat{D}_t$ be estimates constructed from $\hat{\beta}_t$. (i) Let $y_{t+1}$ be predicted as $\hat{C}_t\hat{D}_tZ_{t+1}$, $\tilde{f} = P^{-1}\Sigma[y_{t+1} - \hat{C}_t\hat{D}_tZ_{t+1}]$. Let $v_t = y_t - C^{-1}DZ_t$ be the reduced form disturbances. Then $P^{1/2}\tilde{f} A N(0, S_{ff})$, $S_{ff} = \Sigma_{j=0}^{\infty}Ev_tv_{t-j}'$. (ii) Suppose that $Z_t = (1, y_{t-1})'$ so that the reduced form of the model is $y_t = C^{-1}DZ_t + C^{-1}u_t = c + \Phi y_{t-1} + C^{-1}u_t$. Let $\Phi_t$ and $\hat{\beta}_t$ be constructed from $\hat{C}_t\hat{D}_t$. Let $y_{t+r}$ be predicted as $(1 + \Phi + \ldots + \Phi^{-1})^t c + \Phi y_t$. $\tilde{f} = P^{-1}\Sigma[y_{t+r} - (1 + \Phi + \ldots + \Phi^{-1})^t c + \Phi y_t]$. Let $v_{t+r} = y_{t+r} - [(1 + \Phi + \ldots + \Phi^{-1}) c + \Phi y_t]$. Then $P^{1/2}\tilde{f} A N(0, S_{ff})$, $S_{ff} = \Sigma_{j=0}^{\infty}Ev_tv_{t-j}'$.

Technical conditions: Let $h_t = u_t \otimes Z_t$, where $Z_t = [z_{1t}]$ is a $(r \times 1)$ vector of instruments. For $B$ defined in assumption 2, suppose that $\beta_t$ is estimated in such a fashion that $B = -(A[\partial h_s(\beta^*)/\partial \beta])^{-1}A$, $A = [-\partial h_s(\beta^*)/\partial \beta]'W$ for some symmetric p.d. $(r \times r)$ matrix $W$. (See line (3) in Table II; see Hansen (1982) for
primitive conditions that ensure that such an estimator obeys the section 3
assumptions.) Suppose that each equation contains a constant term whose
coefficient is unrestricted (i.e., \( g(X_t, \beta^*) = c + g_1(X_{1t}, \beta_1) \) with \( \partial g_1/\partial c = 0 \)), that \( Z_t \)
contains a constant term, that for some \( n \) (possibly \( n = \infty \)) one or more of the
following hold: (i) \( \omega^{-1} = \sum_{j=-n}^{n} h_{t-j} \omega_{t-j}' \) and for all \( j \) and for any \( i, s, \) and \( m \) such
that \( 1 \leq i \leq r, 1 \leq s, m \leq I, \) \( E_{zt} u_{zt} u_{mt-j} = E_{zt} u_{zt} u_{mt-j} \), or (ii) \( \omega^{-1} = \sum_{j=-n}^{n} [(E_{zt} u_{zt}') \Phi(E_{zt} Z_t')] \),
or (iii) \( k = I r \). (Note that heteroskedasticity of \( u_t \) conditional on cross-products
of stochastic regressors is allowed by (i).)

IV. Additional references

White, Halbert, 1984, Asymptotic Theory for Econometricians, New York: Academic
Press.