Consistent GMM Residuals-Based Tests of Functional Form
Appendix B: Omitted Proofs and Simulation Results

Jonathan B. Hill[1]

Dept. of Economics, University of North Carolina

May 30, 2011

In this appendix we characterize the data generating process and required limits under Assumption 3 (Section B.1), the proof of Theorem A.1 (Section B.2), and omitted simulation results (Section B.3). Citations used only here are referenced at the end.

B.1 ASSUMPTION 3 Let $\delta > 0$ be a small number, and let vec(z) denote the vector that stacks columns of z. Write $\Gamma^{(*)} = \Gamma \times \cdots \times \Gamma$ the space of matrices $\xi^{(-)} = [\xi^{(-)}_{(1)}, ..., \xi^{(-)}_{(k)}]$, $\xi^{(+)} = [\xi^{(+)}_{(1)}, ..., \xi^{(+)}_{(k)}]$ or $[\xi^{(-)}, \xi^{(+)}]$ (the dimension of $\Gamma^{(*)}$ depends on the case). Let $\hat{\xi}$ denote either $\hat{\xi}^{(-)}$, $\hat{\xi}^{(+)}$ or $[\hat{\xi}^{(-)}, \hat{\xi}^{(+)}]$. Similarly, $\mathfrak{Z} = \mathfrak{Z}_1 \otimes \cdots \otimes \mathfrak{Z}_k$ where $\mathfrak{Z}_i$ denote either $\mathfrak{Z}_i^{(+)} := \{\text{argsup}_{\gamma \in \Gamma}\{(\partial/\partial\gamma_i)E[\epsilon_t F(\gamma' x_{t-1})]\}\}$, $\mathfrak{Z}_i^{(-)} := \{\text{arginf}_{\gamma \in \Gamma}\{(\partial/\partial\gamma_i)E[\epsilon_t F(\gamma' x_{t-1})]\}\}$ or $\mathfrak{Z}_i^{(-)} \otimes \mathfrak{Z}_i^{(+)}$.

3.1: *Each* $z_t \in \{y_t, \tilde{x}_{1,t}, ..., \tilde{x}_{k-1,t}\}$ *is* L_2-Near Epoch Dependent [NED] *on a strong mixing base* $\{\epsilon_t\}$ *with mixing coefficients* $\alpha_i = O(i^{-\lambda})$ *for some* $\lambda > 1$*, in the following uniform sense (see Hill 2008, cf. Gallant and White 1988):*

$$\sup_{t \in \mathbb{Z}} \left\| z_t - E\left[z_t | \{\epsilon_{t-i}\}_{i=0}^m\right] \right\|_r = O\left(v(m)\right) \text{ where } v(m) \to 0 \text{ as } m \to \infty.$$

The error ϵ_t *has an almost everywhere positive continuous density, and* $||\epsilon_t||_{4+\delta} < \infty$.

Remark: We require $\{y_t, \tilde{x}_t\}$ to be L_2-NED on a strong mixing base $\{\epsilon_t\}$ in order to exploit weak limit theory in Hill (2008), cf. Bierens (1991). Mixing errors allow for GARCH and stochastic volatility errors (e.g. Carrasco and Chen 2002), any strong mixing $\{y_t, \tilde{x}_t\}$ is automatically L_2-NED on itself, and in general NED captures a broad array of linear and nonlinear time series with geometric or hyperbolic memory decay. Examples include threshold-type models (e.g. An and Huang 1996), and various nonlinear AR-GARCH (e.g. Meitz and Saikkonen 2008). See Hill (2011b) for a variety of examples and references in other contexts.

3.2: *The parameter space* Φ *is a compact subset of* $\mathbb{R}^p$. *The known response function* $f_t(\phi)$ *is* $\Im_{t-1}$-*measurable, almost surely twice continuously differentiable on* Φ.

3.3:
i. Estimating equations are $m_t(\phi) = (y_t - f_t(\phi)) \times c_t(\phi)$, $c_t : \Phi \to \mathbb{R}^q$, $q \geq p$, $c_t(\phi)$ *is* $\Im_{t-1}$-*measurable and almost surely differentiable on* Θ. *Further* $E[m_t(\phi)] = 0$ if and only if $\phi = \phi_0$, *a unique interior point of* Φ*, hence* ϵ_t *is covariance orthogonal to* $c_t(\phi_0)$ *under either hypothesis. The GMM estimator is* $\hat{\phi} = \arg\inf_{\phi \in \Phi}\{\bar{m}(\phi)'\hat{\Upsilon}\bar{m}(\phi)\}$ *where* $\bar{m}(\phi) = 1/n \sum_{t=1}^n m_t(\phi)$, $\hat{\Upsilon} \in \mathbb{R}^{q \times q}$ *is positive definite, and* $|\hat{\Upsilon} - \Upsilon| \xrightarrow{p} 0$ *for some positive definite* $\Upsilon \in \mathbb{R}^{q \times q}$.
ii. $\hat{A}(\phi, \gamma, \xi) \xrightarrow{p} A(\phi, \gamma, \xi) \in \mathbb{R}^{(k+1) \times q}$ *uniformly on* $\Phi \times \Gamma \times \Gamma^{(*)}$*, where* $A(\phi, \gamma, \xi) =$

[1] Dept. of Economics, University of North Carolina-Chapel Hill, www.unc.edu/∼jbhill, jbhill@email.unc.edu.

$-\operatorname{plim}_{n\to\infty}\{(1/n\sum_{t=1}^{n} w_t(\gamma,\xi)(\partial/\partial\phi')f_t(\phi)) \times (H(\phi))^{-1}(\partial/\partial\phi')\bar{m}(\phi) \times \Upsilon)\}$ *and* $H(\phi) = -\operatorname{plim}_{n\to\infty}\{(\partial/\partial\phi')\bar{m}(\phi) \times \Upsilon \times (\partial/\partial\phi)\bar{m}(\phi)\}$.

iii. There exists a unique element ξ^* *of* $\mathfrak{Z}$ *satisfying* $|\operatorname{vec}(\hat{\xi}) - \operatorname{vec}(\xi^*)| = O_p(1/\sqrt{n})$. *Further,*

$$\sup_{\gamma\in\Gamma}\sup_{1\le t\le n}\left|w_t(\gamma,\hat{\xi}) - w_t(\gamma,\xi^*)\right| = O_p(1/\sqrt{n}) \text{ and } \sup_{\phi\in\Phi,\gamma\in\Gamma}\left|\hat{A}(\phi,\gamma,\hat{\xi}) - A(\phi,\gamma,\xi^*)\right| = O_p(1/\sqrt{n})$$

$$\sup_{\gamma\in\Gamma}\sup_{1\le t\le n}\left|\hat{g}_t(\gamma,\hat{\xi}) - g_t(\gamma,\xi^*)\right| = O_p(1/\sqrt{n}) \text{ and } \sup_{\phi\in\Phi}|H_n(\phi) - H(\phi)| = O_p(1/\sqrt{n}).$$

3.4:

i. There exists a mapping $\mu : \Gamma \times \Gamma^{(*)} \to \mathbb{R}$ *satisfying* $(1/n)\sum_{t=1}^{n} u_t g_t(\gamma,\xi) \xrightarrow{p} \lim_{n\to\infty}(1/n)\sum_{t=1}^{n} E[u_t g_t(\gamma,\xi)] = \mu(\gamma,\xi)$ *uniformly on* $\Gamma \times \Gamma^{(*)}$.

ii. There exists a mapping $\Sigma : \Gamma \times \Gamma \to \mathbb{R}^{(k+1\times(k+1)}$ *satisfying* $(1/n)\sum_{t=1}^{n} E[\epsilon_t^2|\Im_{t-1}] \times g_t(\gamma_1,\xi)g_t(\gamma_2,\xi)' \xrightarrow{p} \Sigma(\gamma_1,\gamma_2)$, $(1/n)\sum_{t=1}^{n}\epsilon_t^2 g_t(\gamma_1,\xi)g_t(\gamma_2,\xi)' \xrightarrow{p} \Sigma(\gamma_1,\gamma_2)$ *and* $(1/n)\sum_{t=1}^{n} E[\epsilon_t^2 g_t(\gamma_1,\xi)g_t(\gamma_2,\xi)'] \to \Sigma(\gamma_1,\gamma_2)$ *uniformly on* $\Gamma \times \Gamma$.

iii. $\limsup_{n\to\infty}\sup_{\gamma\in\Gamma} 1/n\sum_{t=1}^{n} E|\epsilon_t^2 u_t g_t(\gamma,\xi)|^{2+\delta} < \infty$.

3.5: *Uniformly on* $1 \le t \le n$: $||\sup_{\gamma\in\Gamma,\xi\in\Gamma^{(*)}}|g_t(\gamma,\xi)|\,||_{4+\delta} < K$, $||\sup_{\gamma\in\Gamma,\xi\in\Gamma^{(*)}}|(\partial/\partial\gamma)g_t(\gamma,\xi)|\,||_{4+\delta} < K$, $||\sup_{\phi\in\Phi}|c_t(\phi)|\,||_{4+\delta} < K$, *and* $||\sup_{\phi\in\Phi}|(\partial/\partial\phi)c_t(\phi)|\,||_{4+\delta} < C$.

3.6: *The local alternative random variable* u_t *is* $\Im_{t-1}$*-measurable, governed by a non-generate distribution.*

B.2 PROOF OF THEOREM A.1

Recall the claim.

THEOREM A.1. *Assume Assumptions 1 and 2, and* $P(E[\epsilon_t|\Im_{t-1}] = 0) < 1$ *hold. Then* $S = \{0\}$ if and only if ϵ_t *is* L_2*-orthogonal to the closed linear span of* $\{x_{i,t}F'(\xi_{(i)}'x_t)\}_{i=1}^{k}$, *and otherwise* S *is empty.*

We only treat $\xi_{(i)} = \xi_{(i)}^{(+)}$, and for the sake of convention assume

$$\frac{\partial}{\partial\gamma_i}E\left[\epsilon_t F(\gamma'x_t)\right]|_{\gamma=\xi_{(i)}} \ge 0.$$

All subsequent results carry over to the general case $(\partial/\partial\gamma_i)E[\epsilon_t F(\gamma'x_t)]|_{\gamma=\xi_{(i)}} \gtreqless 0$, and to $\operatorname{arginf}_{\gamma\in\Gamma}\{(\partial/\partial\gamma_i)E[\epsilon_t F(\gamma'x_t)]\}$.

The proof requires one supporting result. Define

$$\varpi(\gamma,\xi) := E\left[\epsilon_t\left(F(\gamma'x_t) - \sum_{i=1}^{k}\gamma_i x_{i,t}F'\left(\xi_{(i)}'x_t\right)\right)\right].$$

LEMMA B.1. *Under Assumptions 1 and 2 if* $P(E[\epsilon_t|\Im_{t-1}] = 0) < 1$ *then* $\varpi(\gamma,\xi) = 0$ *for* $\gamma \in \Gamma$ *if and only if* $\gamma = 0$.

Remark: The two structures $E[\epsilon_t F(\gamma'x_t)]$ and $\sum_{i=1}^{k}\gamma_i E[\epsilon_t x_{i,t}F'(\xi_{(i)}'x_t)]$ are equal only at the origin $\gamma = 0$. But this means if $E[\epsilon_t F(\gamma'x_t)] = 0$ under H_1 for any $\gamma \ne 0$ then $\sum_{i=1}^{k}\gamma_i E[\epsilon_t x_{i,t}F'(\xi_{(i)}'x_t)] \ne 0$ hence at least one moment condition $E[\epsilon_t x_{i,t}F'(\xi_{(i)}'x_t)] \ne 0$. Similarly, if all moments $E[\epsilon_t x_{i,t}F'(\xi_{(i)}'x_t)] = 0$ then $E[\epsilon_t F(\gamma'x_t)] \ne 0$ for all $\gamma \ne 0$.

PROOF OF THEOREM A.1. Assume $P[E(\epsilon_t|\Im_{t-1}) = 0] < 1$. By Lemma B.1 we know for each $\gamma \ne 0$

$$\varpi(\gamma,\xi) = E\left[\epsilon_t F(\gamma'x_t)\right] - \sum_{i=1}^{k}\gamma_i E\left[\epsilon_t x_{i,t}F'\left(\xi_{(i)}'x_t\right)\right] \ne 0.$$

Trivially, therefore, at least one moment condition $E\left[\epsilon_t F(\gamma' x_t)\right]$, $E[\epsilon_t x_{1,t} F'(\xi'_{(1)} x_t)]$, ..., or $E[\epsilon_t x_{k,t} F'(\xi'_{(k)} x_t)]$ must be non-zero, hence $E[\epsilon_t w_t(\gamma, \xi)] \neq 0$ for every $\gamma \neq 0$.

Finally, under Assumption 1

$$E[[\epsilon_t w_t(0, \xi)] = \left[0, E\left[\epsilon_t x_{1,t} F'\left(\xi'_{(1)} x_t\right)\right], ..., E\left[\epsilon_t x_{k,t} F'\left(\xi'_{(k)} x_t\right)\right]\right]',$$

hence $E[\epsilon_t w_t(0, \xi)] = 0$ *if and only if* ϵ_t is orthogonal to the closed linear span of $\{x_{i,t} F'(\xi'_{(i)} x_t)\}_{i=1}^k$. $\mathcal{QED}$.

In order to prove Lemma B.1 we require an easy extension of Theorem 1 of Bierens and Ploberger (1997) and Theorem 2.3 and Corollary 3.9 of Stinchcombe and White (1998). Let $h_t : D \to \mathbb{R}^k$ be an $\Im_{t-1}$-measurable, uniformly bounded function, where D is an arbitrary subset of $\mathbb{R}^l$ for some $l \geq 0$. Write $h_t = h_t(\delta)$ by convention when $l = 0$. Examples of $h_t(\delta)$ include x_t, $|x_t|^\delta \times sign(x_t)$ and $(\partial/\partial\phi) f_t(\phi)$ where $\delta = \phi$ provided x_t and $(\partial/\partial\phi) f_t(\phi)$ are bounded with probability one.

LEMMA B.2. *Let weight F satisfy Assumption 1. If $P(E[\epsilon_t|\Im_{t-1}] = 0) < 1$ then for each $\delta \in D$ the set $S = \bigcap_{i=1}^k \{\gamma \in \mathbb{R}^k : E[\epsilon_t h_{i,t}(\delta) F(\gamma' x_t)] = 0\}$ and $P(\gamma' x_t \in R_0) = 1\}$, has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^k$.*

Remark: By Assumption 1 and Corollary 3.9 of Stinchcombe and White (1998), Lemma B.2 holds with $F(\cdot)$ replaced by $F'(\cdot)$. If x_t is not bounded with probability one then replace it with any Borel measurable, bounded one-to-one mapping $\Psi(x_t)$ (Bierens 1990).

PROOF OF LEMMA B.1. Recall

$$\xi_{(i)} \in \mathfrak{Z}_i := \left\{\underset{\gamma\in\Gamma}{\text{argsup}} \left\{\frac{\partial}{\partial \gamma_i} \mu(\gamma)\right\}\right\} = \left\{\underset{\gamma\in\Gamma}{\text{argsup}} \left\{E\left[\epsilon_t \frac{\partial}{\partial \gamma_i} F(\gamma' x_t)\right]\right\}\right\}.$$

The construction of $\xi_{(i)}$ implies for all $\gamma \in \Gamma$

$$E\left[\epsilon_t x_{i,t} F'(\gamma' x_t)\right] \leq E\left[\epsilon_t x_{i,t} F'\left(\xi'_{(i)} x_t\right)\right] = \sup_{\gamma \in \Gamma} E\left[\epsilon_t x_{i,t} F'(\gamma' x_t)\right] \tag{1}$$

and Assumptions 1 and 3 imply $\varpi(0, \xi) = E\left[\epsilon_t F(0' x_t)\right] = 0$. Differentiate $\varpi(\gamma, \xi)$ with respect to γ_j, and add and subtract $E[\epsilon_t x_{j,t} F'(0' x_t)]$:

$$\begin{aligned} \frac{\partial}{\partial \gamma_j} \varpi(\gamma, \xi) &= E\left[\epsilon_t x_{j,t} \left\{F'(\gamma' x_t) - F'(0' x_t)\right\}\right] \\ &\quad - E\left[\epsilon_t x_{j,t} \{F'\left(\xi'_{(j)} x_t\right) - F'(0' x_t)\}\right] \leq 0. \end{aligned} \tag{2}$$

Thus $\varpi(\gamma, \xi)$ is zero at $\gamma = 0$ and is weakly decreasing in γ, and (1) implies $E[\epsilon_t x_{j,t} \{F'(\xi'_{(j)} x_t) - F'(0' x_t)\}] \geq 0$ $\forall j = 1...k$.

In order to prove weak inequality (2) is in fact strict, there are two cases.

Case 1 ($E[\epsilon_t x_{j,t} \{F'(\xi'_{(j)} x_t) - F'(0' x_t)\}] = 0$): Trivially

$$E\left[\epsilon_t x_{j,t} \{F'(\gamma' x_t) - F'(0' x_t)\}\right]|_{\gamma=0} = 0.$$

Lemma B.2 therefore implies there exists an open neighborhood $N(0) \subset \Gamma$ of zero satisfying

$$E\left[\epsilon_t x_{j,t} \{F'(\gamma' x_t) - F'(0' x_t)\}\right] \neq 0 \quad \forall \gamma \in N(0)/0.$$

Since by assumption $E[\epsilon_t x_{j,t} \{F'(\xi'_{(j)} x_t) - F'(0' x_t)\}] = 0$ we deduce from (2)

$$E\left[\epsilon_t x_{j,t} \{F'(\gamma' x_t) - F'(0' x_t)\}\right] < 0 \quad \forall \gamma \in N(0)/0.$$

Thus $\varpi(\gamma, \xi)$ is zero at $\gamma = 0$, strictly decreasing arbitrarily close to $\gamma = 0$, and weakly decreasing everywhere else. This implies $\varpi(\gamma, \xi) \neq 0$ for every $\gamma \neq 0$.

Case 2 ($E[\epsilon_t x_{j,t}\{F'(\xi'_{(j)}x_t) - F'(0'x_t)\}] > 0$): Use (2) to deduce for each $j = 1...k$

$$\frac{\partial}{\partial \gamma_j}\varpi(\gamma, \xi)|_{\gamma=0} = 0 - E\left[\epsilon_t x_{j,t}\left\{F'\left(\xi'_{(j)}x_t\right) - F'\left(0'x_t\right)\right\}\right] < 0.$$

Again, $\varpi(\gamma, \xi)$ is zero at $\gamma = 0$, strictly decreasing at $\gamma = 0$ and weakly decreasing everywhere else. $\mathcal{QED}$.

B.3 OMITTED SIMULATION RESULTS Finally, we present omitted simulation results for fixed CM tests.

Table B.1: Fixed CM Tests Rejection Frequencies[a]

n = 200					
	AR	LSTAR	ESTAR	SETAR	BILIN
L-$\mathcal{T}_n(\gamma_0)$	.055[b]	.079	.749	.455	.058
E-$\mathcal{T}_n(\gamma_0)$	.054	.205	.546	.279	.035
L-$\mathcal{B}_n(\gamma_0)$[c]	.039	.080	.115	.115	.060
E-$\mathcal{B}_n(\gamma_0)$	.049	.109	.197	.153	.031
L-$\mathcal{H}_n(\gamma_*)$[d]	.054	.061	.087	.132	.050
E-$\mathcal{H}_n(\gamma_*)$	.042	.085	.174	.151	.025
UMP(γ_0)[e]	-	.987	1.00	1.00	.299
n = 500					
	AR	LSTAR	ESTAR	SETAR	BILIN
L-$\mathcal{T}_n(\gamma_0)$	.046	.370	1.00	.913	.128
E-$\mathcal{T}_n(\gamma_0)$	.049	.691	.949	.827	.041
L-$\mathcal{B}_n(\gamma_0)$	.057	.212	.428	.357	.061
E-$\mathcal{B}_n(\gamma_0)$	.046	.290	.632	.454	.018
L-$\mathcal{H}_n(\gamma_*)$	.039	.215	.433	.339	.062
E-$\mathcal{H}_n(\gamma_*)$	.051	.299	.647	.442	.015
UMP(γ_0)	-	1.00	1.00	1.00	.483
n = 2000					
	AR	LSTAR	ESTAR	SETAR	BILIN
L-$\mathcal{T}_n(\gamma_0)$	.045	.968	1.00	1.00	.419
E-$\mathcal{T}_n(\gamma_0)$	.042	1.00	1.00	1.00	.029
L-$\mathcal{B}_n(\gamma_0)$	.053	.742	.967	.891	.075
E-$\mathcal{B}_n(\gamma_0)$	.053	.893	1.00	.974	.005
L-$\mathcal{H}_n(\gamma_*)$	.062	.732	.933	.901	.064
E-$\mathcal{H}_n(\gamma_*)$	.061	.891	.984	.974	.004
UMP(γ_0)	-	1.00	1.00	1.00	.751

a. The nuisance parameter is fixed at the mid-point $\gamma_0 = [5.25, ..., 5.25]'$ of $\Gamma = [.5, 10]^{p+1}$.
b. L = logistic; E = exponential. Values are rejection frequency at the 5% level.
c. Bierens' (1990) and Lee et al's (1996) CM test with logistic or exponential weight $F(\gamma'\psi_t)$.
d. Hill's (2008) STAR test with logistic or exponential weight $x_t F(\gamma'\psi_t)$.
e. The Uniformly Most Powerful test. Each test statistic is designed to be UMP for the particular H_1.
In simulations not reported here, each statistic obtains empirical size roughly equal to nominal size.

REFERENCES

Bierens H. J. (1991). Least Squares Estimation of Linear and Nonlinear ARMAX Models under Data Dependence, *Annales d'Economie et de Statistique* 20/21, 143-169.

Hill, J.B. (2011b). Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics, *Econometric Theory*: forthcoming.