Econometrica Supplementary Material

SUPPLEMENT TO "NEARLY OPTIMAL TESTS WHEN A NUISANCE PARAMETER IS PRESENT UNDER THE NULL HYPOTHESIS"

(*Econometrica*, Vol. 83, No. 2, March 2015, 771–811)

BY GRAHAM ELLIOTT, ULRICH K. MÜLLER, AND MARK W. WATSON

APPENDIX B: DETAILS OF THE ALGORITHM USED TO COMPUTE THE POWER BOUNDS IN SECTION 4.3

SIMILARLY TO THE DISCUSSION IN SECTION 3, discretize $\Theta_{1,S}$ by defining M_1 base distributions $\Psi_{1,i}$ with support in $\Theta_{1,S}$, and denote $f_{1,i} = \int f_\theta \, d\Psi_{i,1}$. The constraint $\inf_{\theta \in \Theta_{1,S}} [\int \varphi f_\theta \, d\nu - \pi_S(\theta)] \geq 0$ on φ thus implies $\int \varphi f_{1,i} \, d\nu \geq \int \tilde{\varphi} f_{1,i} \, d\nu$, $i = 1, \ldots, M_1$. For notational consistency, denote the discretization of Θ_0 by $f_{0,i}$, $i = 1, \ldots, M_0$. Let $\mu = (\mu_0', \mu_1')' \in \mathbb{R}^{M_0} \times \mathbb{R}^{M_1}$, and consider tests of the form

$$\varphi_\mu = \mathbf{1}\left[g + \sum_{i=1}^{M_1} \exp(\mu_{1,i}) f_{1,i} > \sum_{i=1}^{M_0} \exp(\mu_{0,i}) f_{0,i} \right].$$

The algorithm is similar to the one described in Section A.2.1, but based on the iterations

$$\mu_{0,j}^{(i+1)} = \mu_{0,j}^{(i)} + \omega\left(\int \varphi_{\mu^{(i)}} f_{0,j} \, d\nu - \alpha \right), \quad j = 1, \ldots, M_0,$$

$$\mu_{1,j}^{(i+1)} = \mu_{1,j}^{(i)} - \omega\left(\int \varphi_{\mu^{(i)}} f_{1,j} \, d\nu - \int \tilde{\varphi} f_{1,i} \, d\nu \right), \quad j = 1, \ldots, M_1.$$

More explicitly, the importance sampling estimators for $\int \varphi_{\mu^{(i)}} f_{0,j} \, d\nu$ and $\int \varphi_{\mu^{(i)}} f_{1,j} \, d\nu$ are given by

$$\begin{aligned} \widehat{\mathrm{RP}}_{0,j}(\mu) = (M_0 N_0)^{-1} \sum_{k=1}^{M_0} \sum_{l=1}^{N_0} \frac{f_{0,j}(Y_{k,l}^0)}{\bar{f}_0(Y_{k,l}^0)} \mathbf{1}\Bigg[& g(Y_{k,l}^0) \\ & + \sum_{i=1}^{M_1} \exp(\mu_{1,i}) f_{1,i}(Y_{k,l}^0) > \sum_{i=1}^{M_0} \exp(\mu_{0,i}) f_{0,i}(Y_{k,l}^0) \Bigg], \\ \widehat{\mathrm{RP}}_{1,j}(\mu) = (M_1 N_0)^{-1} \sum_{k=1}^{M_1} \sum_{l=1}^{N_0} \frac{f_{1,j}(Y_{k,l}^1)}{\bar{f}_1(Y_{k,l}^1)} \mathbf{1}\Bigg[& g(Y_{k,l}^1) \\ & + \sum_{i=1}^{M_1} \exp(\mu_{1,i}) f_{1,i}(Y_{k,l}^1) > \sum_{i=1}^{M_0} \exp(\mu_{0,i}) f_{0,i}(Y_{k,l}^1) \Bigg], \end{aligned}$$

 DOI: 10.3982/ECTA10535

where $\bar{f}_0(y) = M_0^{-1}\sum_{j=1}^{M_0} f_{0,j}(y)$ and $\bar{f}_1(y) = M_1^{-1}\sum_{j=1}^{M_1} f_{1,j}(y)$, and $Y_{k,l}^0$ and $Y_{k,l}^1$ are N_0 i.i.d. draws from density $f_{0,k}$ and $f_{1,k}$, respectively. For future reference, for two given points $\hat{\Lambda}_0^* = (\hat{\lambda}_{0,1}^*, \dots, \hat{\lambda}_{0,M_0}^*)$ and $\hat{\Lambda}_1^* = (\hat{\lambda}_{1,1}^*, \dots, \hat{\lambda}_{1,M_1}^*)$ in the M_0- and M_1-dimensional simplex, respectively, define

$$\begin{aligned}
\widehat{\mathrm{RP}}_{0,j}(\mathrm{cv}_0, \mathrm{cv}_1) &= (M_0 N_0)^{-1} \sum_{k=1}^{M_0} \sum_{l=1}^{N_0} \frac{f_{0,j}(Y_{k,l}^0)}{\bar{f}_0(Y_{k,l}^0)} \mathbf{1}\Bigg[g(Y_{k,l}^0) \\
&\qquad + \mathrm{cv}_1 \sum_{i=1}^{M_1} \hat{\lambda}_{1,i}^* f_{1,i}(Y_{k,l}^0) > \mathrm{cv}_0 \sum_{i=1}^{M_0} \hat{\lambda}_{0,i}^* f_{0,i}(Y_{k,l}^0) \Bigg], \\
\widehat{\mathrm{RP}}_{1,j}(\mathrm{cv}_0, \mathrm{cv}_1) &= (M_1 N_0)^{-1} \sum_{k=1}^{M_1} \sum_{l=1}^{N_0} \frac{f_{1,j}(Y_{k,l}^1)}{\bar{f}_1(Y_{k,l}^1)} \mathbf{1}\Bigg[g(Y_{k,l}^1) \\
&\qquad + \mathrm{cv}_1 \sum_{i=1}^{M_1} \hat{\lambda}_{1,i}^* f_{1,i}(Y_{k,l}^1) > \mathrm{cv}_0 \sum_{i=1}^{M_0} \hat{\lambda}_{0,i}^* f_{0,i}(Y_{k,l}^1) \Bigg], \\
\widehat{\mathrm{RP}}_{g}(\mathrm{cv}_0, \mathrm{cv}_1) &= N_1^{-1} \sum_{l=1}^{N_1} \mathbf{1}\Bigg[g(Y_l) + \mathrm{cv}_1 \sum_{i=1}^{M_1} \hat{\lambda}_{1,i}^* f_{1,i}(Y_l) \\
&\qquad > \mathrm{cv}_0 \sum_{i=1}^{M_0} \hat{\lambda}_{0,i}^* f_{0,i}(Y_l) \Bigg],
\end{aligned}$$

where Y_l are N_1 i.i.d. draws from density g. The algorithm now proceeds in the following steps:

1. For each k, $k = 1, \dots, M_0$, generate N_0 i.i.d. draws $Y_{k,l}^0$, $l = 1, \dots, N_0$, with density $f_{0,k}$, and for each $k = 1, \dots, M_1$, generate N_0 i.i.d. draws $Y_{k,l}^1$, $l = 1, \dots, N_0$, with density $f_{1,k}$. The draws $Y_{k,l}^0$ and $Y_{k,l}^1$ are independent across k and l.
2. Compute and store $g(Y_{k,l})$, $f_{0,j}(Y_{k,l}^0)$, $\bar{f}_0(Y_{k,l}^0)$, $j, k = 1, \dots, M_0$, $l = 1, \dots, N_0$, as well as $f_{1j}(Y_{k,l}^1)$ and $\bar{f}_1(Y_{k,l}^1)$, $j, k = 1, \dots, M_1$, $l = 1, \dots, N_0$.
3. Compute the (estimated) power $\pi_j \approx \int \tilde{\varphi} f_{1,j}\, d\nu$ of $\tilde{\varphi} = \chi \varphi_S$ under $f_{1,j}$ via $\pi_j = (M_1 N_0)^{-1} \sum_{k=1}^{M_1} \sum_{l=1}^{N_0} \frac{f_{1,j}(Y_{k,l}^1)}{\bar{f}_1(Y_{k,l}^1)} \chi(Y_{k,l}^1) \varphi_S(Y_{k,l}^1)$, $j = 1, \dots, M_1$.
4. Set $\mu^{(0)} = (-2, \dots, -2) \in \mathbb{R}^{M_0+M_1}$.
5. Compute $\mu^{(i+1)}$ from $\mu^{(i)}$ via $\mu_{0,j}^{(i+1)} = \mu_{0,j}^{(i)} + \omega(\widehat{\mathrm{RP}}_{0,}(\mu^{(i)}) - \alpha)$, $j = 1, \dots, M_0$ and $\mu_{1,j}^{(i+1)} = \mu_{1,j}^{(i)} - \omega(\widehat{\mathrm{RP}}_{1,j}(\mu^{(i)}) - \pi_j)$, $j = 1, \dots, M_1$ with $\omega = 2$, and repeat this step $O = 600$ times. Denote the resulting elements in the M_0- and M_1-dimensional simplex by $\hat{\Lambda}_0^* = (\hat{\lambda}_{0,1}^*, \dots, \hat{\lambda}_{0,M_0}^*)$ and $\hat{\Lambda}_1^* = (\hat{\lambda}_{1,1}^*, \dots, \hat{\lambda}_{1,M_1}^*)$, where $\hat{\lambda}_{0,j}^* = \exp(\mu_{0,j}^{(O)}) / \sum_{k=1}^{M_0} \exp(\mu_{0,k}^{(O)})$ and $\hat{\lambda}_{1,j}^* = \exp(\mu_{1,j}^{(O)}) / \sum_{k=1}^{M_1} \exp(\mu_{1,k}^{(O)})$.

6. Compute the number $\text{cv}^*_{0,0}$ such that the test $\mathbf{1}[g > \text{cv}^*_{0,0}\sum_{i=1}^{M_0}\hat{\lambda}^*_{0,i}f_{0,i}]$ is exactly of (Monte Carlo) level α under the mixture $\sum_{j=1}^{M_0}\hat{\lambda}^*_{0,j}f_{0,j}$, that is, solve $\sum_{j=1}^{M_0}\hat{\lambda}^*_{0,j}\widehat{\text{RP}}_{0,j}(\text{cv}_{0,0},0)=\alpha$ for $\text{cv}_{0,0}$. If the resulting test has power under the mixture $\sum_{j=1}^{M_1}\hat{\lambda}^*_{1,j}f_{1,j}$ larger than $\sum_{j=1}^{M_1}\hat{\lambda}^*_{1,j}\pi_j$, that is, if $\sum_{j=1}^{M_1}\hat{\lambda}^*_{1,j}(\widehat{\text{RP}}_{1,j}(\text{cv}^*_{0,0},0)-\pi_j)\geq 0$, then the power constraint does not bind, and the power bound is given by $\widehat{\text{RP}}_g(\text{cv}^*_{0,0},0)$.

7. Otherwise, compute the two numbers cv^*_0 and cv^*_1 such that the test $\mathbf{1}[g+\text{cv}^*_1\sum_{i=1}^{M_1}\hat{\lambda}^*_{1,i}f_{1,i} > \text{cv}^*_0\sum_{i=1}^{M_0}\hat{\lambda}^*_{0,i}f_{0,i}]$ is of (Monte Carlo) level α under the mixture $\sum_{j=1}^{M_0}\hat{\lambda}^*_{0,j}f_{0,j}$, and of power equal to $\sum_{j=1}^{M_1}\hat{\lambda}^*_{1,j}\pi_j$ under the mixture $\sum_{j=1}^{M_1}\hat{\lambda}^*_{1,j}f_{1,j}$, that is, solve the two equations $\sum_{j=1}^{M_0}\hat{\lambda}^*_{0,j}\widehat{\text{RP}}_{0,j}(\text{cv}_0,\text{cv}_1)=\alpha$ and $\sum_{j=1}^{M_1}\hat{\lambda}^*_{1,j}\widehat{\text{RP}}_{1,j}(\text{cv}_0,\text{cv}_1)=\sum_{j=1}^{M_1}\hat{\lambda}^*_{1,j}\pi_j$ for $(\text{cv}_0,\text{cv}_1)\in\mathbb{R}^2$. The power bound is then given by $\widehat{\text{RP}}_g(\text{cv}^*_0,\text{cv}^*_1)$.

APPENDIX C: ADDITIONAL DETAILS FOR THE APPLICATIONS

The following lemma is useful for obtaining closed form expressions in many of the applications.

LEMMA 6: *For* $c>0$, $\int_{-\infty}^{a}\exp[sd-\frac{1}{2}s^2c^2]\,ds=\sqrt{2\pi}c^{-1}\exp[\frac{1}{2}d^2/c^2]\Phi(ac-d/c)$, *where* Φ *is the c.d.f. of a standard normal.*

PROOF: Follows from "completing the square." *Q.E.D.*

In all applications, the M base distributions on Θ_0 are either uniform distributions, or point masses. Size control is always checked by computing the Monte Carlo rejection probability at all δ that are end or mid-points of these intervals, or that are simple averages of the adjacent locations of point masses, respectively (this check is successful in all applications). The power bound calculations under the power constraint of Section 4.3 use the same $M_0=M$ base distributions under the null, and the M_1 base distributions with support on $\Theta_{1,S}$ all set β to the same value as employed in F, and use the same type of base distribution on δ as employed in the discretization of Θ_0.

C.1. *Running Example*

The base distributions on Θ_0 are the uniform distributions on the intervals $\{[0,0.04],[0,0.5],[0.5,1],[1,1.5],\dots,[12,12.5]\}$. The base distributions on $\Theta_{1,S}$ have $\beta\in\{-2,2\}$ and δ uniform on the intervals $\{[9,9.5],[9.5,10],\dots,[13,13.5]\}$.

C.2. *Behrens–Fisher*

Limit Experiment and Standard Best Test

We analyze convergence as $\delta \to \infty$, that is, as $\sigma_2/\sigma_1 \to 0$. The convergence as $\delta \to -\infty$ follows by the same argument.

Consider the four-dimensional observation $\tilde{Y} = (\bar{x}_1, \bar{x}_2, s_1, s_2)$, with density

$$\frac{\sqrt{n_1 n_2}}{\sigma_1^2 \sigma_2^2} \phi\left(\frac{\bar{x}_1 - \mu_1}{\sigma_1/\sqrt{n_1}}\right) \phi\left(\frac{\bar{x}_2 - \mu_2}{\sigma_2/\sqrt{n_2}}\right) f_{n_1}\left(\frac{s_1}{\sigma_1}\right) f_{n_2}\left(\frac{s_2}{\sigma_2}\right),$$

where ϕ is the density of a standard normal, and f_n is the density of a chi-distributed random variable with $n-1$ degrees of freedom, divided by $\sqrt{n-1}$. Now set $\mu_2 = 0$, so that $b = \beta = (\mu_1 - \mu_2)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ implies $\mu_1 = b\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$. Also, set $\sigma_1 = \exp(d)$ and $\sigma_2 = \exp(-\Delta_n)$, so that $\delta = \log(\sigma_1/\sigma_2) = \Delta_n + d$. This is without loss of generality as long as one restricts attention to tests that are invariant to the transformations described in the main text.

Let $f_{n,h}$ be the density of $\tilde{Y}$ in this parameterization, where $h = (b, d)$. Further, let $f_{X,h}$ be the density of the bivariate vector $X = (X_b, X_d)$ where $X_b \sim \mathcal{N}(b\exp(d), \exp(2d))$ and X_d is an independently chi-distributed random variable with $n-1$ degrees of freedom, divided by $\sqrt{n-1}$. With $\tilde{Y}$ distributed according to $f_{n,0}$, and X distributed according to $f_{X,0}$, we find, for any finite set $H \subset \mathbb{R}^2$,

$$\begin{aligned}
&\left\{\frac{f_{n,h}(\tilde{Y})}{f_{n,0}(\tilde{Y})}\right\}_{h\in H} \\
&\quad = \left\{\left(\exp(2d)\phi\left(\frac{\bar{x}_1 - b\sqrt{\exp(2d)/n_1 + \exp(-2\Delta_n)/n_2}}{\exp(d)/\sqrt{n_1}}\right)\right.\right. \\
&\qquad \left.\left.\times f_{n_1}\left(\frac{s_1}{\exp(d)}\right)\right) \Big/ \left(\phi\left(\frac{\bar{x}_1}{1/\sqrt{n_1}}\right) f_{n_1}(s_1)\right)\right\}_{h\in H} \\
&\quad \Rightarrow \left\{\frac{\exp(2d)\phi\left(\frac{X_b - b\exp(d)/\sqrt{n_1}}{\exp(d)/\sqrt{n_1}}\right) f_{n_1}\left(\frac{X_d}{\exp(d)}\right)}{\phi\left(\frac{X_b}{1/\sqrt{n_1}}\right) f_{n_1}(X_d)}\right\}_{h\in H} \\
&\quad = \left\{\frac{f_{X,h}(X)}{f_{X,0}(X)}\right\}_{h\in H},
\end{aligned}$$

so that Condition 1 is satisfied. Thus, tests of $H_0 : b = 0$ against $H_1 : b \neq 0$ based on X form an upper bound on the asymptotic power as $\Delta_n \to \infty$ of invariant tests based on $\tilde{Y}$. The standard (and admissible) test $\varphi_S^{\lim}$ based on X is the usual t-test $\mathbf{1}[|X_b|/X_d > \text{cv}]$. Further, a straightforward calculation shows that the invariant test $\varphi_S = \mathbf{1}[|\frac{\bar{x}_1-\bar{x}_2}{\sqrt{s_1^2/n_1+s_2^2/n_2}}| > \text{cv}] = \mathbf{1}[|Y_\beta| > \text{cv}]$ has the same asymptotic rejection probability as $\varphi_S^{\lim}$ for all fixed values of h.

Computational Details

It is computationally convenient to consider the one-to-one transformation $(t, r) = ((\bar{x}_1 - \bar{x}_2)/s_2, s_1/s_2) = (\sqrt{\frac{e^{2Y_\delta}}{n_1} + \frac{1}{n_2}} Y_\beta, e^{Y_\delta})$ with parameters $\eta = \mu_1 - \mu_2 = \beta\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ and $\omega = \sigma_1/\sigma_2 = \exp(\delta)$. A transformation of variable calculation shows that the density of (t, r) is given by

$$
\begin{aligned}
f(t, r) = {}& \frac{(n_1-1)^{n_1/2}(n_2-1)^{n_2/2}\omega}{r^2\Gamma\left(\frac{1+n_1}{2}\right)\Gamma\left(\frac{1+n_2}{2}\right)}\sqrt{\frac{n_1 n_2}{\pi\left(n_1+\omega^2 n_2\right)}}\left(\frac{r}{\omega}\right)^{n_1} \\
& \times 2^{(1-n_1-n_2)/2}\exp\left[-\frac{1}{2}\frac{\eta^2 n_1 n_2}{n_1+n_2\omega^2}\right] \\
& \times \int_0^\infty s^{n_1+n_2-2}\exp\Big[\Big(2\eta n_1 n_2 st - s^2\Big((n_2-1)n_2\omega^4 + n_1^2 r^2 \\
& - n_2\omega^2 r^2 + n_1\left(n_2\omega^2\left(1+r^2+t^2\right) - \omega^2 - r^2\right)\Big)/\omega^2\Big) \\
& / \left(2\left(n_1+n_2\omega^2\right)\right)\Big]\, ds
\end{aligned}
$$

where Γ denotes the Gamma function. The integral is recognized as being proportional to the (n_1+n_2-2)th absolute moment of a half normal. In particular, for $c > 0$, $\int_0^\infty \exp[-\frac{1}{2}s^2c^2]s^n\, ds = 2^{(n-1)/2}\Gamma(\frac{1+n}{2})c^{-(n+1)}$, and following Dhrymes (2005),

$$
\int_0^\infty \exp\left[sd - \frac{1}{2}s^2c^2\right]s^n\, ds = \exp\left[\frac{1}{2}\frac{d^2}{c^2}\right]\frac{d^n}{c^{2n+1}}\sum_{l=0}^{n}\binom{n}{l}\left(-\frac{c}{d}\right)^l I_l\left(\frac{d}{c}\right),
$$

where

$$
\begin{aligned}
I_l(h) &= \int_{-\infty}^{h}\exp\left[-\frac{1}{2}z^2\right]z^l\, dz \\
&= \begin{cases} 2^{(l-1)/2}\left(\left(1+(-1)^l\right)\Gamma\left(\frac{1+l}{2}\right) - \tilde{\Gamma}\left(\frac{1+l}{2}, \frac{h^2}{2}\right)\right) \\ \quad \text{for } h > 0, \\ 2^{(l-1)/2}(-1)^l\tilde{\Gamma}\left(\frac{1+l}{2}, \frac{h^2}{2}\right) \\ \quad \text{for } h \le 0, \end{cases}
\end{aligned}
$$

with $\tilde{\Gamma}$ the upper incomplete Gamma function, $\tilde{\Gamma}(a,x)=\int_x^\infty s^{a-1}e^{-s}\,ds$.

The base distributions on Θ_0 are uniform distributions for δ on the intervals $\{[-12.5,-12],[-12,-11.5],\ldots,[12,12.5]\}$, and the base distributions on $\Theta_{1,S}=\{(\beta,\delta):|\delta|>9\}$ have δ uniform on $\{[-14,-13.5],[-13.5,-13],\ldots,[-9.5,-9]\}\cup\{[9,9.5],[9.5,10],\ldots,[14.5,15]\}$. The corresponding integrals are computed via Gaussian quadrature using 10 nodes (for this purpose, the integral under the alternative is split up in intervals of length 2). For $n_1=n_2$, symmetry around zero is imposed in the calculation of the ALFD.

C.3. *Break Date*

Wiener processes are approximated with 1000 steps. Symmetry around zero is imposed in the calculation of the ALFD, and the set of base distribution for $|\delta|$ contains uniform distributions on $\{[0,1],[1,2],\ldots,[19,20]\}$.

C.4. *Predictive Regression*

Limit Experiment

As in the main text, let $\delta=r_\delta(\Delta_n,d)=\Delta_n-\sqrt{2\Delta_n}d$ and $\beta=r_\beta(\Delta_n,\beta)=\sqrt{2\Delta_n/(1-\rho^2)}b$. After some algebra, under $h=0$,

$$\ln\frac{f_{n,h}(G)}{f_{n,0}(G)}=\sqrt{2\Delta_n}\left(\frac{(b-d\rho)\int_0^1 W^\mu_{x,\Delta_n}(s)\,dW_y(s)}{\sqrt{1-\rho^2}}+d\int_0^1 W_{x,\Delta_n}(s)\,dW_x(s)\right)-\Delta_n\left(d^2\int_0^1 W_{x,\Delta_n}(s)^2\,ds+\frac{(b-d\rho)^2}{1-\rho^2}\int_0^1 W^\mu_{x,\Delta_n}(s)^2\,ds\right).$$

Now suppose the following convergence holds as $\Delta_n\to\infty$:

$$\begin{pmatrix}\sqrt{2\Delta_n}\int_0^1 W_{x,\Delta_n}(s)\,dW_x(s)\\ \sqrt{2\Delta_n}\int_0^1 W^\mu_{x,\Delta_n}(s)\,dW_y(s)\\ 2\Delta_n\int_0^1 W_{x,\Delta_n}(s)^2\,ds\\ 2\Delta_n\int_0^1 W^\mu_{x,\Delta_n}(s)^2\,ds\end{pmatrix}\Rightarrow\begin{pmatrix}Z_x\\ Z_y\\ 1\\ 1\end{pmatrix}, \tag{32}$$

where Z_x and Z_y are independent $\mathcal{N}(0,1)$. Then, as $\Delta_n \to \infty$,

$$\begin{aligned}
&\ln \frac{f_{n,h}(G)}{f_{n,0}(G)} \\
&\quad \Rightarrow -\frac{1}{2}\frac{b^2 - 2bd\rho + d^2 - 2(b-\rho d)\sqrt{1-\rho^2}Z_y - 2d(1-\rho^2)Z_x}{1-\rho^2} \\
&\quad = \begin{pmatrix} X_b \\ X_d \end{pmatrix}' \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} b \\ d \end{pmatrix} - \frac{1}{2}\begin{pmatrix} b \\ d \end{pmatrix}' \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} b \\ d \end{pmatrix},
\end{aligned}$$

where $X_b = \rho Z_x + \sqrt{1-\rho^2}Z_y$ and $X_d = Z_x$, and Condition 1 follows from the continuous mapping theorem.

To establish (32), note that

$$\begin{aligned}
&\int_0^1 W_{x,\Delta_n}^{\mu}(s)^2\,ds = \int_0^1 W_{x,\Delta_n}(s)^2\,ds - \left(\int_0^1 W_{x,\Delta_n}(s)\,ds\right)^2, \\
&\int_0^1 W_{x,\Delta_n}(s)\,dW_x(s) = \frac{1}{2}\big(W_{x,\Delta_n}(1)^2 - 1\big) + \Delta_n \int W_{x,\delta}(s)^2\,ds, \\
&\int_0^1 W_{x,\Delta_n}^{\mu}(s)\,dW_y(s) = \int_0^1 W_{x,\Delta_n}(s)\,dW_y(s) - W_y(1)\int_0^1 W_{x,\Delta_n}(s)\,ds.
\end{aligned}$$

Thus, with $t = (t_1, \dots, t_4)$ and $\mathrm{i} = \sqrt{-1}$,

$$\phi_n(t) = E\left[\exp\left[\mathrm{i}t'\begin{pmatrix} \sqrt{2\Delta_n}\int_0^1 W_{x,\Delta_n}(s)\,dW_x(s) \\ \sqrt{2\Delta_n}\int_0^1 W_{x,\Delta_n}^{\mu}(s)\,dW_y(s) \\ 2\Delta_n \int_0^1 W_{x,\Delta_n}(s)^2\,ds \\ 2\Delta_n \int_0^1 W_{x,\Delta_n}^{\mu}(s)^2\,ds \end{pmatrix}\right]\right]$$

$$= E\left[\exp\left[\mathrm{i}\begin{pmatrix} -\sqrt{2\Delta_n}t_2 \\ \sqrt{2\Delta_n}t_2 \\ 2\Delta_n t_3 + 2\Delta_n t_4 + \sqrt{2}\Delta_n^{3/2}t_1 \\ \sqrt{\Delta_n/2}t_1 \\ -2\Delta_n t_4 \end{pmatrix}'\right.\right.$$

$$\left.\left. \times \begin{pmatrix} W_y(1)\int_0^1 W_{x,\Delta_n}(s)\,ds \\ \int_0^1 W_{x,\Delta_n}(s)\,dW_y(s) \\ \int_0^1 W_{x,\Delta_n}(s)^2\,ds \\ W_{x,\Delta_n}(1)^2 \\ \left(\int_0^1 W_{x,\Delta_n}(s)\,ds\right)^2 \end{pmatrix} - \mathrm{i}t_1\sqrt{\Delta_n/2}\right]\right].$$

Note that

$$E\left[E\left[\exp\left[\mathrm{i}\begin{pmatrix} -\sqrt{2\Delta_n}t_2 \\ \sqrt{2\Delta_n}t_2 \end{pmatrix}'\begin{pmatrix} W_y(1)\int_0^1 W_{x,\Delta_n}(s)\,ds \\ \int_0^1 W_{x,\Delta_n}(s)\,dW_y(s) \end{pmatrix}\right]\middle| W_x\right]\right]$$

$$= E\left[\exp\left[-\frac{1}{2}\begin{pmatrix} -\sqrt{2\Delta_n}t_2 \\ \sqrt{2\Delta_n}t_2 \end{pmatrix}'\right.\right.$$

$$\times \begin{pmatrix} \left(\int_0^1 W_{x,\Delta_n}(s)\,ds\right)^2 & \left(\int_0^1 W_{x,\Delta_n}(s)\,ds\right)^2 \\ \left(\int_0^1 W_{x,\Delta_n}(s)\,ds\right)^2 & \int_0^1 W_{x,\Delta_n}(s)^2\,ds \end{pmatrix}$$

$$\left.\left. \times \begin{pmatrix} -\sqrt{2\Delta_n}t_2 \\ \sqrt{2\Delta_n}t_2 \end{pmatrix}\right]\right].$$

Thus

$$\phi_n(t) = E\left[\exp\left[\begin{pmatrix} 2\Delta_n t_3 \mathrm{i} + 2\Delta_n t_4 \mathrm{i} + \sqrt{2}\Delta_n^{3/2} t_1 \mathrm{i} - \Delta_n t_2^2 \\ \sqrt{\Delta_n/2}\, t_1 \mathrm{i} \\ -2\Delta_n t_4 \mathrm{i} + \Delta_n t_2^2 \end{pmatrix}'\right.\right.$$

$$\left.\left.\times \begin{pmatrix} \int_0^1 W_{x,\Delta_n}(s)^2\,ds \\ W_{x,\Delta_n}(1)^2 \\ \left(\int_0^1 W_{x,\Delta_n}(s)\,ds\right)^2 \end{pmatrix} - \mathrm{i}t_1\sqrt{\Delta_n/2}\right]\right]$$

$$= E\left[\exp\left[\begin{pmatrix} l_{n,1} \\ l_{n,2} \\ l_{n,3} \end{pmatrix}' \begin{pmatrix} \int_0^1 W_{x,\Delta_n}(s)^2\,ds \\ W_{x,\Delta_n}(1)^2 \\ \left(\int_0^1 W_{x,\Delta_n}(s)\,ds\right)^2 \end{pmatrix} - \mathrm{i}t_1\sqrt{\Delta_n/2}\right]\right]$$

$$= \det\bigl(I_2 - 2V(\gamma_n)\Omega_n\bigr)^{-1/2} \exp\left[-\mathrm{i}t_1\sqrt{\Delta_n/2} - \frac{1}{2}(\gamma - \Delta_n)\right],$$

where $\gamma_n = \sqrt{\Delta_n^2 - 2l_{n,1}}$, $\Omega_n = \mathrm{diag}(l_{n,2} + \frac{1}{2}(\gamma_n - \Delta_n), l_{n,3})$, and

$$V(\gamma) = \int \begin{pmatrix} e^{-\gamma(1-s)} \\ \dfrac{1 - e^{-\gamma(1-s)}}{\gamma} \end{pmatrix} \begin{pmatrix} e^{-\gamma(1-s)} \\ \dfrac{1 - e^{-\gamma(1-s)}}{\gamma} \end{pmatrix}' ds,$$

and the third equality applies Lemma 1 of Elliott and Müller (2006). Let $Y_n = \mathrm{diag}(1, \sqrt{\Delta_n})$. A calculation now shows that, as $\Delta_n \to \infty$,

$$\begin{aligned}
&Y_n V(\gamma_n) Y_n \to 0, \\
&Y_n^{-1}\Omega_n Y_n^{-1} = O(1), \\
&-\mathrm{i}t_1\sqrt{\Delta_n/2} - \frac{1}{2}(\gamma_n - \Delta_n) \to -\frac{1}{2}t_1^2 - \frac{1}{2}t_2^2 + t_3\mathrm{i} + t_4\mathrm{i},
\end{aligned}$$

so that $\phi_n(t)$ converges pointwise to the characteristic function of the right hand side of (32), which proves (32).

Computational Details

Ornstein–Uhlenbeck and stochastic integrals are approximated with 1000 steps. The base distributions on Θ_0 are point masses at the points $\delta \in$

$\{0^2, 0.5^2, \dots, 14.25^2\}$, and the base distributions on $\Theta_{1,S}$ are point masses on $\delta \in \{160, 165, \dots, 190\}$, with the corresponding value of β as in (24) with $b = 1.645$.

Modified Version of Campbell and Yogo (2006) Test

In the main text, we compared our test to the Campbell and Yogo (2006) (CY) test for predictive ability of a persistent regressor. As noted there, our test controls size uniformly for $\delta \geq 0$. In contrast, the CY test inverts the DF-GLS unit root test, which, as noted by Mikusheva (2007), results in a confidence interval for the autoregressive parameter r that does not have uniform coverage properties over all δ.

We modified the CY procedure so that the confidence set for r was constructed using pointwise t-tests of $H_0 : r = r_0$ for all possible values of r_0 (as in Hansen (1999)). As in CY, the nominal size for the subsequent (augmented) t-test of $\gamma = 0$ with r known was set at 5%, and the coverage rate for the pointwise confidence sets for r were determined so that the overall size of the test for $\gamma = 0$ was 5%. Figure 8 compares the (asymptotic local) power of this particular modification of CY with the nearly optimal test derived in Section 5.3. As expected, the modified CY test does not show a drop-off in power for large δ. It does show somewhat lower power than the original CY test for moderate values of δ, although this may be a reflection of the particular size correction we employed.

C.5. *Set Identified Parameter*

Limit Experiment

We consider convergence for $\beta \geq 0$ as $\Delta_L \to \infty$; the convergence for $\beta \leq 0$ follows analogously.

Set $\beta = b$, $\delta_P = d_P$, and $\delta_L = \Delta_n + d_L$, so that in this parameterization, $\mu_l = \tau(b, d_P) = \mathbf{1}[b > 0]b - \mathbf{1}[b = 0]d_P$ and $\mu_u = \Delta_n + d_L + \tau(b, d_P)$. For any fixed $h = (b, d_L, d_P) \in \mathbb{R}^2 \times [0, \infty)$, as $\Delta_n \to \infty$,

$$\log \frac{f_{n,h}(Y)}{f_{n,0}(Y)} = \begin{pmatrix} Y_l \\ Y_u - \Delta_n \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} \tau(b, d_P) \\ \tau(b, d_P) + d_L \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \tau(b, d_P) \\ \tau(b, d_P) + d_L \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} \tau(b, d_P) \\ \tau(b, d_P) + d_L \end{pmatrix}.$$

Because $(Y_l, Y_u - \Delta_n)' \sim \mathcal{N}(0, \Sigma)$ for $h = 0$ as $\Delta_n \to \infty$, Theorem 9.4 in van der Vaart (1998) implies that Condition 1 holds with

$$X = \begin{pmatrix} X_b \\ X_d \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \tau(b, d_P) \\ \tau(b, d_P) + d_L \end{pmatrix}, \Sigma \right).$$

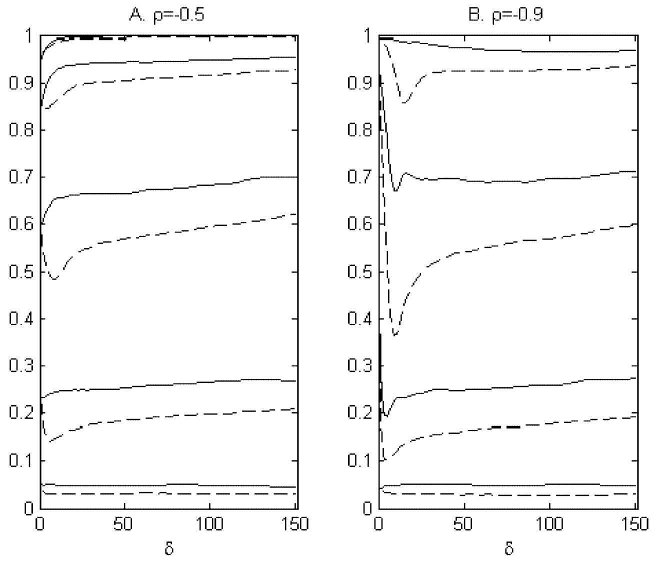

FIGURE 8.—Power comparison with modified Campbell and Yogo (2006) test. Dashed lines are the power of the modified CY test, and solid lines are the power of the nearly optimal tests $\varphi^{\varepsilon}_{\Lambda^*,S,\chi}$ of Section 5.3 for $b \in \{0, 1, 2, 3, 4\}$.

The test of $H_0 : b = 0$, $(d_L, d_P) \in \mathbb{R} \times [0, \infty)$ against $H_1 : b > 0$, $d_L \in \mathbb{R}$ in this limiting experiment thus corresponds to $H_0 : E[X_b] \leq 0$ against $H_1 : E[X_b] > 0$, with $E[X_d]$ unrestricted under both hypotheses. The uniformly best test is thus given by $\varphi^{\lim}_S(x) = \mathbf{1}[x_b > \text{cv}]$: This follows by the analytical least favorable distribution result employed below (29) assuming $d_P = 0$ known, and since $\varphi^{\lim}_S$ is of level α also for $d_P > 0$, putting all mass at $d_P = 0$ is also least favorable in this more general testing problem.

A test with the same asymptotic rejection probability for any fixed h is given by $\varphi_S(y) = \mathbf{1}[y_l > \text{cv}]$.

Computational Details

The base distributions on Θ_0 have δ_L uniform on the intervals $\{[0, 0.1], [0, 0.5], [0.5, 1], [1, 1.5], \ldots, [12.5, 13]\}$, with δ_P an equal probability mixture on the two points $\{0, \delta_L\}$. The base distributions on $\Theta_{1,S}$ have δ_L uniform on the intervals $\{[9, 9.25], [9.25, 9.5], \ldots, [11.75, 12]\}$.

C.6. *Regressor Selection*

Symmetry around zero is imposed in the computation of the ALFD. The base distributions on Θ_0 are point masses at $|\delta| \in \{0, 0.2, 0.4, \ldots, 9\}$.

REFERENCES

CAMPBELL, J. Y., AND M. YOGO (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60. [10,11]

DHRYMES, P. J. (2005): "Moments of Truncated (Normal) Distributions," Working Paper, Columbia University. [5]

ELLIOTT, G., AND U. K. MÜLLER (2006): "Minimizing the Impact of the Initial Condition on Testing for Unit Roots," *Journal of Econometrics*, 135, 285–310. [9]

HANSEN, B. E. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81, 594–607. [10]

MIKUSHEVA, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, 75, 1411–1452. [10]

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge, U.K.: Cambridge University Press. [10]

University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, U.S.A.; grelliott@ucsd.edu,

Dept. of Economics, Princeton University, Princeton, NJ 08544, U.S.A.; umueller@princeton.edu,

and

Dept. of Economics, Princeton University, Princeton, NJ 08544, U.S.A. and NBER; mwatson@princeton.edu.

Manuscript received January, 2012; final revision received September, 2014.