Diebold, F.X., Ohanian, L. and Berkowitz, J. (1998),
"Dynamic Equilibrium Economies: A Framework for Comparing Models and Data,"
Review of Economic Studies, 65, 433-452.

Dynamic Equilibrium Economies: A Framework for Comparing Models and Data

Francis X. Diebold
University of Pennsylvania
and
NBER

Lee E. Ohanian
University of Minnesota
and
University of Pennsylvania

Jeremy Berkowitz
Federal Reserve Board

Revised March 1997

Address correspondence to:

Francis X. Diebold
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297

Abstract: We propose a constructive, multivariate framework for assessing agreement between (generally misspecified) dynamic equilibrium models and data, which enables a complete second-order comparison of the dynamic properties of models and data. We use bootstrap algorithms to evaluate the significance of deviations between models and data, and we use goodness-of-fit criteria to produce estimators that optimize economically-relevant loss functions. We provide a detailed illustrative application to modeling the U.S. cattle cycle.

Acknowledgments: The Co-Editor and referees provided helpful and constructive input, as did participants at meetings of the Econometric Society, the CEPR, the NBER, and numerous university seminars. We gratefully acknowledge additional help from Bill Brown, Fabio Canova, Tim Cogley, Bob Lucas, Ellen McGrattan, Danny Quah, Lucrezia Reichlin, Sherwin Rosen, Chris Sims, Tony Smith, Jim Stock, Mark Watson, and especially Lars Hansen, Adrian Pagan, and Tom Sargent. All remaining errors and inaccuracies are ours. José Lopez provided dedicated research assistance in the early stages of this project. We thank the National Science Foundation, the Sloan Foundation and the University of Pennsylvania Research Foundation for support.

1. Introduction

Dynamic equilibrium models are now used routinely in many fields. Such models, for example, have been used to address a variety of macroeconomic issues, including business-cycle fluctuations, economic growth, and the effects of government policies.[1] Additional prominent fields of application include international economics, public economics, industrial organization, labor economics, and agricultural economics.[2]

At present, however, many important questions regarding the empirical implementation of dynamic equilibrium models remain incompletely answered. The questions fall roughly into two methodological groups. The first group involves issues related to assessing model adequacy, and the second involves issues related to model estimation. We contribute to an emerging literature that has begun to deal with both issues, including Watson (1993), King and Watson (1992, 1996), Canova, Finn and Pagan (1994), Kim and Pagan (1994), Pagan (1994), Leeper and Sims (1994), Cogley and Nason (1995), and Hansen, McGrattan and Sargent (1997). A 1996 *Journal of Economic Perspectives* symposium focused on these issues, and two important messages emerged:[3] (1) dynamic equilibrium models, like all models, are intentionally simple abstractions and therefore should not be

---

[1] Among many others, see Kydland and Prescott (1982), Hansen (1985), Christiano and Eichenbaum (1995), and Rotemberg and Woodford (1996) (business cycles), Lucas (1988), Jones and Manuelli (1990), Rebelo (1991), and Greenwood, Hercowitz, and Krusell (1997) (growth), and Lucas (1990), Cooley and Hansen (1992), and Ohanian (1997) (policy effects).

[2] Among many others, see Backus, Kehoe and Kydland (1994) (international economics), Auerbach and Kotlikoff (1987) (public economics), Ericson and Pakes (1995) (industrial organization), Rust (1989) (labor economics), and Rosen, Murphy and Scheinkman (1994) (agricultural economics).

[3] See Kydland and Prescott (1996), Sims (1996) and Hansen and Heckman (1996).

construed as the true data generating process, and (2) formal methods should be developed and used to help us assess the models more thoroughly. In this paper, we take a step in that direction.

Some parts of our framework are new, while others build on earlier work in interesting ways. In many respects, our work begins where Watson (1993) ends. With an eye toward future research, Watson notes that "... one of the most informative diagnostics ... is the plot of the model and data spectra," and he recommends that in the future researchers "present both model and data spectra as a convenient way of comparing their complete set of second moments."[4] Our methods, which are based on comparison of model and data spectral density functions, can be used to assess the performance of a model (for a given set of parameters), to estimate model parameters, and to test hypotheses about parameters or models. To elaborate, our approach is:

A. Frequency-domain and multivariate. Working in the frequency domain enables decomposition of variation across frequencies, which is often useful, and the multivariate focus facilitates simple examination of cross-variable correlations and lead-lag relationships, at the frequencies of interest.

B. Based on a full second-order comparison of model and data dynamics. This is in contrast to a common approach used in the business cycle literature of comparing only a few variances and covariances of detrended variables from the model economy and the actual economy. The spectrum provides a complete summary of Gaussian time series dynamics and an approximate summary of non-Gaussian time series dynamics.

C. Based on the realistic assumption that all models are misspecified. We regard all of the models we entertain as false, in which case traditional statistical methods lose some of their appeal.

---

[4] He also notes that his failure to study cross-variable relationships is a potentially important omission.

D. Graphical and constructive. The framework permits one to assess visually and quickly the dimensions along which a model performs well, and the dimensions along which it performs poorly.

E. Based on a common set of tools that can be used by researchers with potentially very different objectives and research strategies. The framework can be used to evaluate strictly calibrated models, and it can also be used formally to estimate and test models.

F. Designed to facilitate statistical inference about objects estimated from data, including spectra, goodness-of-fit measures, model parameters, and test statistics. Bootstrap methods play an important role in that regard; we develop and use a simple nonparametric bootstrap algorithm.

G. Mathematically convenient. Under regularity conditions, the spectrum is a bounded continuous function, which makes for convenient mathematical developments.

All of the classical ideas of business-cycle analysis discussed, for example, by Lucas (1977) have spectral analogs, ranging from univariate persistence (typical spectral shape) to multivariate issues of comovement (coherence) and lead-lag relationships (phase shifts) at business-cycle frequencies. We highlight these links and draw upon the business-cycle literature for motivation in the methodological sections 2 and 3. The methods we develop, however, are not wed to macroeconomics in any way; rather, they can be used in a variety of fields. Therefore, to introduce researchers in different areas to the use of our framework, we apply our methods to a simple and accessible, yet rich, microeconomic model in section 4. We conclude in section 5.

2. Assessing Agreement Between Model and Data

Our basic strategy is to assess models by comparing model spectra to data spectra. Our goal is provision of a graphical framework that facilitates visual comparisons of model spectra to interval estimates of data spectra. We compute model spectra exactly (either

analytically or numerically); thus, they have no sampling uncertainty. Sampling error does, however, affect the sample data spectra, which are of course just estimates of true but unknown (population) data spectra. We exploit well-established procedures for estimating spectra, and we develop and use bootstrap techniques to assess the sampling uncertainty of estimated spectra.[5]

2a. Estimating Spectra

Consider the N-variate linearly regular covariance stationary stochastic process,

$$y_t = \mu + B(L)\,\varepsilon_t = \mu + \sum_{i=-\infty}^{\infty} B_i\,\varepsilon_{t-i}$$

$$E(\varepsilon_t \varepsilon_s') = \begin{cases} \Omega & \text{if } t = s \\ 0 & \text{otherwise,} \end{cases}$$

where $E(\varepsilon_t) = 0$, $B_0 = I$, and the coefficients are square summable (in the matrix sense).[6] The autocovariance function is $\Gamma(\tau) = \sum_{i=-\infty}^{\infty} B_i\,\Omega\,B_{i+\tau}'$ and the spectral density function is $F(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma(\tau)\, e^{-i\omega\tau}$, $-\pi<\omega<\pi$.

Consider now a generic off-diagonal element of $F(\omega)$, $f_{kl}(\omega)$. In polar form, the cross-spectral density is $f_{kl}(\omega) = ga_{kl}(\omega) \exp[i\, ph_{kl}(\omega)]$, where $ga_{kl}(\omega) = [re^2(f_{kl}(\omega)) + im^2(f_{kl}(\omega))]^{1/2}$ is the gain or amplitude, and where $ph_{kl}(\omega) = \arctan\{im(f_{kl}(\omega)) / re(f_{kl}(\omega))\}$ is the phase. As is

[5] Alternatively, one could fix the data spectrum, and assess sampling error in the model spectrum by simulating repeated realizations from the model. The two approaches are essentially complementary, corresponding to the "Wald" and "Lagrange multiplier" testing perspectives. See, for example, Gregory and Smith (1991).

[6] In many cases, detrending of some sort will be necessary to achieve covariance stationarity.

well known, the gain tells how the amplitude of y_l is multiplied in contributing to the amplitude of y_k at frequency ω, and phase measures the lead of y_k over y_l at frequency ω. (The phase shift in time units is $ph(\omega)/\omega$.) We shall often find it convenient to examine coherence rather than gain, where the coherence is defined as $coh_{kl}(\omega) = \frac{ga_{kl}^2(\omega)}{f_{kk}(\omega)\ f_{ll}(\omega)}$, which measures the squared correlation between y_k and y_l at frequency ω.

Given a sample path $\{y_{1t}, ..., y_{Nt}\}_{t=1}^{T}$, we estimate the Nx1 mean vector μ with $\bar{y} = (\bar{y}_1, ..., \bar{y}_N)'$. From this point onward, we assume that all sample paths have been centered around this sample mean. We estimate the autocovariance function with $\hat{\Gamma}(\tau) = [\hat{\gamma}_{kl}(\tau)]$ $(k = 1, ..., N, l = 1, ..., N)$, where $\hat{\gamma}_{kl}(\tau) = \frac{1}{T}\sum_{\tau=1}^{T-|\tau|} y_{kt}y_{l,t+\tau}$, $\tau = 0, \pm1, ..., \pm(T-1)$. We estimate the spectral density matrix using the Blackman-Tukey lag-window approach in which we replace the sample spectral density function, $\hat{F}(\omega_j) = \frac{1}{2\pi}\sum_{\tau=-(T-1)}^{(T-1)} \hat{\Gamma}(\tau)e^{-i\omega_j\tau}$ $(\omega_j = \frac{2\pi j}{T},\ j = 1, ..., \frac{T}{2}-1)$ with one involving the "windowed" sample autocovariance sequence, $F^*(\omega_j) = \frac{1}{2\pi}\sum_{\tau=-(T-1)}^{(T-1)} \Lambda(\tau)\ \hat{\Gamma}(\tau)\ e^{-i\omega\tau}$, where $\Lambda(\tau)$ is a matrix of lag windows. The Blackman-Tukey procedure results in a consistent estimator if we adjust the lag window $\Lambda(\tau)$ with sample size in such a way that variance and bias decline simultaneously.[7] We then obtain the sample coherence and phase at any frequency ω_j by transforming the appropriate elements of $F^*(\omega_j)$.

2b. Assessing Sampling Variability

A key issue for our purposes is how to ascertain the sampling variability of the estimated spectral density function. To do so, we use an algorithm for resampling from time

[7] Alternatively, of course, one may smooth the sample spectral density function directly. The duality between the two approaches, for appropriate window choices, is well known. See Priestley (1981).

series data, which we call the Cholesky factor bootstrap.[8] The basic idea is straightforward. First we compute the Cholesky factor of the sample covariance matrix of the series of interest. We then exploit the fact that, up to second order, the series of interest can be written as the product of the Cholesky factor and serially uncorrelated disturbances, which can be easily bootstrapped using parametric or non-parametric procedures.[9] An important feature of this very simple approach is that it can be used to bootstrap objects other than the spectral density function. Later, for example, we will use it to assess the uncertainty in a model's estimated parameters.

First we need some definitions and notation. Let $z_t = (y_{1t}, ..., y_{Nt})'$, and let $z = (z_1', z_2', ..., z_T')'$. Then $z \sim (1 \otimes \mu, \Sigma)$, where 1 is an N-dimensional column vector of ones, and $\Sigma = \text{Toeplitz}(\Gamma(0), \Gamma(1), ..., \Gamma(T-1))$. By symmetry and positive definiteness, we can write $\Sigma = PP'$, where the unique Cholesky factor P is lower triangular. We estimate Σ by $\hat{\Sigma} = \text{Toeplitz}(\hat{\Gamma}(0), \hat{\Gamma}(1), ..., \hat{\Gamma}(T-1))$, where $\hat{\Gamma}(\tau) = \frac{1}{T} \sum_{t=1}^{T-|\tau|} z_t \, z_{t+|\tau|}'$, $\tau = 0, \pm 1, ..., \pm(T-1)$; this ensures that we can write $\hat{\Sigma} = \hat{P}\hat{P}'$, where the unique Cholesky factor $\hat{P}$ is lower triangular. Now let $\{\lambda_{|i-j|}\}_{|i-j|=0}^{T-1}$ be a set of decreasing weights applied to

---

[8] The Cholesky factor bootstrap is closely related to the Ramos (1988) bootstrap. We develop the Cholesky factor bootstrap in the time domain, however, whereas Ramos proceeds in the frequency domain.

[9] Note that the Cholesky factor bootstrap will miss nonlinear dynamics such as GARCH -- it is designed to capture only second-order dynamics, in identical fashion to standard (as opposed to higher-order) spectral analysis. Users should be cautious in employing our procedure if nonlinearities are suspected to be operative, as would likely be the case, for example, for high-frequency financial data. Such nonlinearities are not likely to be as important for the lower-frequency data typically analyzed in many areas of macroeconomics, public finance, international economics, industrial organization, agricultural economics, etc.

the successive off-diagonal blocks of $\hat{\Sigma}$, and call the resulting matrix Σ^*. Finally, let P^* be the Cholesky factor of Σ^*.

The fact that $z \sim (1 \otimes \mu, PP')$ implies that data generated by drawing $\varepsilon^{(i)} \overset{iid}{\sim} (0, I_{NT})$ and forming

$$z^{(i)} = \mu_z + P\varepsilon^{(i)},$$

where $\mu_z = 1 \otimes \mu$, will have the same second-order properties as the observed data. In practice we replace the unknown population first and second moments with the consistent estimates described above. Thus, to perform a parametric bootstrap, we draw $\varepsilon^{(i)} \sim N(0, I_{NT})$, form

$$z^{(i)} = \bar{z} + P^*\varepsilon^{(i)} \sim N(\bar{z}, \Sigma^*),$$

where $\bar{z} = 1 \otimes \bar{y}$, and then compute both the estimates $F^{*(i)}(\omega_j)$, $j = 1, ..., \frac{T}{2}-1$, $i = 1, ..., R$ and confidence intervals. Alternatively, to perform a nonparametric bootstrap, we note that $\varepsilon^{(i)} = P^{-1}(z - \mu_z)$. In practice, we draw $\varepsilon^{(i)}$ with replacement from $P^{*-1}(z - \bar{z})$, form

$$z^{(i)} = \bar{z} + P^*\varepsilon^{(i)} \sim (\bar{z}, \Sigma^*),$$

from which we compute $F^{*(i)}(\omega_j)$, $j = 1, ..., \frac{T}{2}-1$, $i = 1, ..., R$, and then construct confidence intervals.

In summary, there are several appealing features of the Cholesky factor bootstrap: (1) it is a very simple procedure, (2) it can be used to bootstrap a variety of objects, (3) it does not involve conditioning on a fitted model and therefore imposes minimal assumptions on dynamics. This last feature may be attractive for researchers who choose not to view the data through the lens of an assumed parametric model. Alternative bootstrap procedures include the VAR bootstrap (e.g., Canova, Finn and Pagan, 1994), which can be a useful approach for

those interested in fitting a specific parametric model to the data. Thus, the Cholesky approach and the VAR approach can be viewed as complementary procedures.

We hasten to add, however, that the literature on bootstrapping time series in general -- and spectra in particular -- is very young and very much unsettled. We still have a great deal to learn about the comparative properties of various bootstraps, both asymptotically and in finite samples, and the conditions required for various properties to obtain. Presently available results differ depending on the specific statistic being bootstrapped, and moreover, only scattered first- and second-order asymptotic results are available, and even less is known about actual finite-sample performance. With this in mind, we present both theoretical and Monte Carlo analyses of the performance of the Cholesky factor bootstrap in two appendixes to this paper. In Appendix 1, we establish first-order asymptotic validity, and in Appendix 2, we document good small-sample performance.

2c. Constructing Confidence Tunnels[10]

If interest centers on only one frequency, we simply use the bootstrap distribution at that frequency to construct the usual bootstrap confidence interval. That is, we find q_T^L, q_T^U such that $P(f^{*(.)}(\omega) \leq q_T^U) = 1-\frac{\alpha}{2}$ and $P(f^{*(.)}(\omega) \geq q_T^L) = 1-\frac{\alpha}{2}$, where $(1-\alpha)$ is the desired confidence level, "L" stands for lower, "U" stands for upper, the "T" subscript indicates that we tailor the band to the finite-sample size T, and the (.) superscript indicates that we take the probability under the bootstrap distribution. The $(1-\alpha)$% two-sided confidence interval is $[q_T^L, q_T^U]$.

[10] In this section, for notational simplicity we focus on confidence tunnels for univariate spectra. As will be clear, the extension to cross spectra is immediate.

However, one often wants to assess the sampling variability of the entire spectral density *function* over many frequencies (e.g., business-cycle frequencies, or perhaps all frequencies) to learn about the broad agreement between data and model. One approach is to form the pointwise bootstrap confidence intervals described above, and then to "connect the dots." But obviously, a set of $(1-\alpha)\%$ confidence intervals constructed for each of n ordinates will not achieve $(1-\alpha)\%$ joint coverage probability. Rather, the actual confidence level will be closer to $(1-\alpha)^n\%$, which holds exactly if the pointwise intervals are independent. A better approach is to use the Bonferroni method to approximate the desired coverage level, by assigning $(1 - \alpha/n)\%$ coverage to each ordinate.[11] The resulting"confidence tunnel" has coverage of at least $(1 - \alpha)\%$ and therefore provides a conservative estimate of the tunnel.[12]

A third approach to confidence tunnel construction is the supremum method of Woodroofe and van Ness (1967) and Swanepoel and van Wyk (1986), which uses an estimate of the (standardized) distribution of $\sup_{0<\omega_j<\pi} |f^*(\omega_j) - f(\omega_j)|$, $\omega_j = \frac{2\pi j}{T}$, $j = 1, ..., \frac{T}{2}-1$, to construct a confidence tunnel for the curve. Specifically,[13]

(1) Calculate $f^{*(.)}(\omega_j)$, $\omega_j = \frac{2\pi j}{T}$, $j = 1, ..., \frac{T}{2}-1$..

[11] In the univariate case, typically n = T/2 - 1. In the multivariate case, the question arises as to "how wide to cast the net" in forming confidence tunnels. One might view each element of the spectral density matrix in isolation, for example, in which case each of the respective confidence tunnels would use n = T/2 -1. At the other extreme, one could use $n = N^2(T/2-1)$, effectively forming a tunnel for the entire matrix.

[12] Bonferroni tunnels achieve the desired coverage only for (1) independent values of the estimated function across ordinates, which is clearly violated in spectral density estimation as the smoothing required for consistency results in averaging across frequencies, and (2) large n, because $(1 - \alpha/n)^n \geq (1 - \alpha)$, for any finite n.

[13] This procedure is similar to the one advocated in Gallant, Rossi and Tauchen (1993).

(2) Find c such that:

$$P\left[\sup_{0<\omega_j<\pi}\left(\frac{\left|f^{*(.)}(\omega_j) - f^{*}(\omega_j)\right|}{\sqrt{2/T}f^{*}(\omega_j)}\right) \leq c\right] = 1-\alpha,$$

where we evaluate the probability with respect to the bootstrap distribution.

(3) Construct the confidence tunnel, $f^{*}(\omega_j) \pm c\sqrt{2/T}f^{*}(\omega_j)$,

$$\omega_j = \frac{2\pi j}{T}, j = 1, ..., \frac{T}{2}-1.$$

Unlike the Bonferroni tunnels, the supremum tunnels attain asymptotically correct coverage rates even with statistical dependence among ordinates. Little is known, however, about the comparative finite-sample performance of the Bonferroni and supremum tunnels, and the supremum tunnels may require very large samples for accurate coverage. [14]

3. Estimation: Maximizing Agreement Between Model and Data

Now we consider estimation, together with the related issues of goodness-of-fit and hypothesis testing. To make the discussion as transparent as possible, we first discuss the univariate case, and then we proceed to the multivariate case.

3a. Univariate

Estimation requires a loss function, or goodness-of-fit measure, for assessing closeness between model and data. A strength of our approach is that many loss functions may be entertained; the particular loss function adopted reflects the user's preferences. [15] In most cases it would seem that a function of the form

---

[14] See Hannan (1970), p. 294.

[15] For an interesting and early discussion of this and related points, see Pagan (1994).

$$C_{gw}(\theta) = \int_0^{\pi} g\left(f_m(\omega;\ \theta),\ f^*(\omega)\right)\ w(\omega)\ d\omega$$

will be adequate. The function g measures the divergence between $f_m(\omega;\ \theta)$ (model spectrum) and $f^*(\omega)$ (estimate of data spectrum).[16] We weight this divergence across frequencies by the function $w(\omega)$. In practice, we replace the integral with a sum over frequencies $\omega_j = \frac{2\pi j}{T}$, $j = 1, \ldots, \frac{T}{2}-1$. Quadratic loss with uniform weighting over all frequencies, for example, corresponds to $g(a, b) = (a-b)^2$ and $w(\omega) = 1$, yielding $C_{gw}(\theta) = \sum_j \left(f_m(\omega_j;\ \theta) - f^*(\omega_j)\right)^2$.

The goodness-of-fit measure may readily be transformed into an estimation criterion by taking

$$\hat{\theta}_{gw} = \underset{\theta}{\text{argmin}}\ C_{gw}(\theta).$$

The Gaussian ML estimator is asymptotically of this form, for a particular and potentially restrictive choice of g, f^*, and w; it is $\underset{\theta}{\text{argmax}}\left(-\frac{1}{2}\sum_j \ln f_m(\omega_j;\ \theta) - \frac{1}{2}\sum_j \frac{\hat{f}(\omega_j)}{f_m(\omega_j;\ \theta)}\right)$.

To compute standard errors and interval estimates for parameters of interest, and to test hypotheses about the elements of $\hat{\theta}_{gw}$, we again use the Cholesky factor bootstrap. We proceed as follows:

(1) At bootstrap replication (i), draw a bootstrap sample of size T using the Cholesky factor algorithm.

(2) Numerically minimize $C_{gw}^{(i)}(\theta)$ to get $\hat{\theta}_{gw}^{(i)}$.

---

[16] Note that the model spectrum is either computable analytically or numerically to any desired degree of accuracy. The data spectrum, on the other hand, is consistently estimable.

(3) Repeat R times.

(4) Compute standard errors, form interval estimates, implement bias corrections, or test hypotheses using the distribution of $\hat{\theta}_{gw}^{(i)}$, i = 1, ..., R.

Note that, unlike most implementations of the bootstrap, ours does not involve conditioning on the model; instead, we generate the bootstrap samples directly from the autocovariance matrix of the data. This is important in our environment, in which all models are best regarded as false.

In closing this section, let us elaborate on our allowance for differential weighting by frequency. There are at least two reasons for entertaining this possibility. First, use of a loss function that weights differentially by frequency may be helpful in dealing with measurement error, which often may not contaminate all frequencies equally. Thus, it would seem prudent to downweight those frequencies that are assumed to be more contaminated by measurement error.

Second, use of a loss function that weights differentially by frequency may be important in misspecified models. For example, as discussed by Hansen and Heckman (1996), model misspecification may contaminate some frequencies more than others. Examples of this include potential contamination at seasonal frequencies, as in the work of Hansen and Sargent (1993) and Sims (1993). Watson (1993) also advocates the use of differential weighting in parameter estimation, for the same reason, although he doesn't pursue the matter. As Watson notes, optimizing a loss function at particular frequencies corresponds to constructing an analog estimator along the lines of Manski (1988).

3b. Multivariate

The multivariate analog of our earlier loss function is

$$C_{GW}(\theta) = \int_0^{\pi} G\left(F_m(\omega;\ \theta),\ F^{*}(\omega)\right) \odot W(\omega)\ d\omega,$$

where $\odot$ denotes component-by-component multiplication. The multivariate analog of our earlier univariate quadratic loss function, for example, is

$C_{GW}(\theta) = \sum_j tr\left(D'(\omega_j;\ \theta)\ D(\omega_j;\ \theta)\right)$, where $D(\omega_j;\ \theta) = F_m(\omega_j;\ \theta) - F^{*}(\omega_j)$, $\omega_j = \frac{2\pi j}{T}$, $j = 1, \ldots, \frac{T}{2}-1$.

The estimation criterion function has the same form as in the univariate case,

$$\hat{\theta}_{GW} = \underset{\theta}{\text{argmin}}\ C_{GW}(\theta),$$

and the bootstrap approaches to computing standard errors, confidence intervals, and hypothesis testing parallel the univariate case precisely. Furthermore, as expected, the multivariate Gaussian ML estimator emerges as a special and potentially restrictive case; it is

$$\underset{\theta}{\text{argmax}} \left(-\frac{1}{2}\sum_j \ln\left|F_m(\omega_j;\ \theta)\right| - \frac{1}{2} tr \sum_j F_m^{-1}(\omega_j;\ \theta)\ \hat{F}(\omega_j) \right).$$

It is worth emphasizing how all parts of the spectrum contribute to loss in the multivariate case. Consider, for example, a bivariate model (variables x and y) under quadratic loss. Then

$$D(\omega_j;\ \theta) = \begin{pmatrix} d_{xx}(\omega_j;\ \theta) & d_{xy}(\omega_j;\ \theta) \\ d_{yx}(\omega_j;\ \theta) & d_{yy}(\omega_j;\ \theta) \end{pmatrix},$$

where

$$d_{xx}(\omega_j;\ \theta) = f_{xx,m}(\omega_j;\ \theta) - f_{xx}^{*}(\omega_j)$$

$$d_{yy}(\omega_j;\ \theta) = f_{yy,m}(\omega_j;\ \theta) - f_{yy}^{*}(\omega_j)$$

$$d_{xy}(\omega_j;\ \theta) = f_{xy,m}(\omega_j;\ \theta) - f^{*}_{xy}(\omega_j)$$

$$d_{yx}(\omega_j;\ \theta) = f_{yx,m}(\omega_j;\ \theta) - f^{*}_{yx}(\omega_j) = \overline{f_{xy,m}(\omega_j;\ \theta)} - \overline{f^{*}_{xy}(\omega_j)} = \overline{d_{xy}}(\omega_j;\ \theta).$$

Thus,

$$\begin{aligned} tr(D'(\omega_j;\ \theta)D(\omega_j;\theta)) &= [d^2_{xx}(\omega_j;\theta) + d_{xy}(\omega_j;\theta)d_{yx}(\omega_j;\theta)] + [d^2_{yy}(\omega_j;\theta) + d_{xy}(\omega_j;\theta)d_{yx}(\omega_j;\theta)] \\ &= d^2_{xx}(\omega_j;\ \theta) + 2|d_{xy}(\omega_j;\ \theta)|^2 + d^2_{yy}(\omega_j;\ \theta) \end{aligned}$$

$$\begin{aligned} = [f_{xx,m}(\omega_j;\ \theta) - f^{*}_{xx}(\omega_j)]^2 &+ 2[re(f_{xy,m}(\omega_j;\ \theta)) - re(f^{*}_{xy}(\omega_j))]^2 \\ &+ 2[im(f_{xy,m}(\omega_j;\ \theta)) - im(f^{*}_{xy}(\omega_j))]^2 + [f_{yy,m}(\omega_j;\ \theta) - f^{*}_{yy}(\omega_j)]^2. \end{aligned}$$

This expression shows clearly how the goodness of fit of both univariate spectra, as well as both the real and imaginary parts of the cross spectrum, contribute to loss.

4. Application: The U.S. Cattle Cycle

Let us begin by summarizing the framework for assessing and estimating dynamic stochastic models developed in sections 2 and 3 of this paper. We first perform a full second-order comparison of model and data by visually comparing model spectra, data spectra, and associated confidence tunnels about the data spectra computed using the simple Cholesky factor bootstrap. To formally assess divergence between model and data spectra, and to estimate model parameters, we develop an explicit loss function that reflects the specific objectives of the investigation. Finally, we assess the sampling distributions of estimated parameters again using the Cholesky factor bootstrap.

It is well known that cattle stock and consumption are among the most periodic time series in economics, with a cycle of roughly ten years in U.S. ("the cattle cycle"). In this

section, we provide a detailed illustration of the use of our assessment and estimation techniques by applying them to an important model of the cattle cycle developed by Rosen, Murphy, and Scheinkman (RMS, 1994). This simple yet rich model allows us to illustrate very clearly the application of all the tools in our framework, and moreover, our findings provide new insight into the RMS model and its agreement with the data.

4a. The Data

We use annual data on U.S. cattle consumption and stock, 1900-1989.[17] We plot the series in Figures 1 and 2, and the cycle is visually apparent. Moreover, the series are clearly trending. Following RMS, we remove a linear trend from each series prior to additional analysis, allowing for a break in the slope of the trend in 1930.[18]

We present the estimated data spectrum in Figure 3.[19] We make use -- here and throughout -- of a matrix graphic with univariate spectra plotted on the main diagonal, coherence in the upper-right corner, and phase in the lower-left corner. Not all frequencies are of equal interest, however. The frequencies most relevant to an investigation of the cattle cycle, typically thought to have a period of roughly ten years, are not those in the entire $[0, \pi]$ range, but rather those in a *subset* that excludes very low and very high frequencies. This presents no problem for our procedures and in fact provides a good opportunity to illustrate the ease with which they can be tailored to study specific applications. Thus, for much of our

[17] The data were kindly supplied by Sherwin Rosen and were originally obtained from *Historical Statistics: Colonial Times to 1970* and *Agricultural Statistics*, published by the U.S. Department of Agriculture.

[18] The fitted trends are also shown in Figures 1 and 2.

[19] We smooth the sample autocovariance function using a Bartlett window with truncation lag 24.

analysis, we concentrate on the frequency band corresponding to periods of 30 years to 4 years, indicated by the shaded region in Figure 3 and subsequent figures.

Four features of the point estimates of the data spectrum stand out. First, the consumption spectrum (and to a lesser extent, the stock spectrum) displays a power concentration at roughly a ten-year cycle. Second, both the consumption and stock spectra otherwise have Granger's (1966) typical spectral shape, with high power at low frequencies, and declining power throughout the frequency range. Third, the coherence between consumption and stock is generally high and varies across frequencies, with a maximum (about .85) at roughly a ten-year cycle. Finally, the phase shift varies with frequency; within the band of interest, the maximum (about one year) is again at roughly a ten year cycle. [20]

In Figure 4 we present the data spectrum along with 90% confidence tunnels computed using the conservative Bonferroni technique in conjunction with the Cholesky-factor bootstrap.[21] To facilitate evaluation, we plot the consumption and stock spectra on a logarithmic scale.[22] All of the point estimates display substantial uncertainty, as manifest in the 90% confidence tunnels. Such uncertainty associated with estimated spectra is typical of economic time series, although it often goes unacknowledged.

4b. The Model

We begin with some accounting identities. The head count of all animals (y_t) is the

[20] Phase shift is measured in years by which consumption leads stock.

[21] The detrended consumption and stock data are nevertheless highly persistent. We present some Monte Carlo evidence in Appendix 2 that indicates that the Cholesky factor bootstrap performs well in such stationary, but highly persistent, environments.

[22] From this point onward, we adopt the log scale for consumption and stock spectra whenever confidence tunnels are included.

sum of the adult breeding stock (x_t), the stock of calves (assumed equal to gx_{t-1}), and the stock of yearlings (assumed equal to gx_{t-2}), where g is a fertility parameter. That is,

$$y_t = x_t + gx_{t-1} + gx_{t-2}.$$

The adult breeding stock consists of surviving stock from the previous period (assumed equal to $(1-\delta)x_{t-1}$) and the yearlings from t-1 entering the adult herd (gx_{t-3}) less the number that are marketed (c_t),

$$x_t = (1-\delta)x_{t-1} + gx_{t-3} - c_t.$$

We are concerned with the equilibrium determination of c_t and y_t. The risk-neutral rancher maximizes the present discounted value of expected profits, which involves equating the expected marginal benefit of marketing an animal for consumption to the expected marginal benefit of holding the animal for breeding. First, suppose that the rancher markets the animal for consumption. He receives net revenue $q_t=p_t-m_t$, where p_t is price and m_t is finishing cost. Alternatively, suppose the rancher holds an animal for breeding. Expected discounted net revenue is the sum of expected discounted revenue from selling tomorrow plus expected discounted revenue from marketing its offspring, less expected total holding costs (z_t), $E_t[\beta(1-\delta)q_{t+1}+\beta^3 gq_{t+3}-z_t]$. Total holding cost equals the sum of time t holding costs (h_t), discounted holding costs of the resultant time t+1 calves, and discounted holding costs of the resultant time t+2 yearlings. That is, $z_t=h_t+\beta g\gamma_0 h_{t+1}+\beta^2 g\gamma_1 h_{t+2}$ (assuming proportional costs for calves and yearlings, γ_0 and γ_1).

In equilibrium, the expected marginal net revenue from marketing for consumption equals the expected marginal discounted net revenue from holding for breeding; that is,

$$E_t[q_t] = E_t[\beta(1-\delta)q_{t+1} + \beta^3 gq_{t+3} - z_t].$$

We close the model by specifying the exogenous processes $\{m_t,\ h_t,\ d_t\}$ as first-order autoregressions.[23] Following RMS, we assume that each of the three shocks has common serial-correlation parameter ρ.

The model structure implies that the reduced-form equations for c_t and y_t can be expressed in terms of a single disturbance, ω_t, which is a linear combination of the independent innovations from the three AR(1) driving processes. In particular, c_t~ARMA(2,1) and y_t~ARMA(4,2):

$$(1-\lambda_1 L)(1-\rho L)\ c_t = -(1 - \phi_1 L)\ \omega_t$$

$$(1-\lambda_1 L)(1-\phi_2 L)(1-\phi_3 L)(1-\rho L)\ y_t = (1 + gL + gL^2)\ \omega_t,$$

where ϕ_1 is the one unstable root and $\{\phi_2,\ \phi_3\}$ are the two stable roots of

$$\phi^3 - (1-\delta)\phi^2 - g = 0,$$

and λ_1 is the one stable root of

$$g\beta^3\lambda^3 + (1-\delta)\beta\lambda - 1 = 0.$$

The associated univariate spectra are

$$f_c(\omega) = \sigma_\omega^2 \frac{|(1 - \phi_1 e^{i\omega})|^2}{|(1 - \lambda_1 e^{i\omega})(1 - \rho e^{i\omega})|^2}$$

$$f_y(\omega) = \sigma_\omega^2 \frac{|(1 + ge^{i\omega} + ge^{2i\omega})|^2}{|(1 - \lambda_1 e^{i\omega})(1 - \phi_2 e^{i\omega})(1 - \phi_3 e^{i\omega})(1 - \rho e^{i\omega})|^2},$$

and the cross spectrum is

$$f_{cy}(\omega) = \frac{-(1 - \phi_1 e^{i\omega})(1 - \phi_2 e^{i\omega})(1 - \phi_3 e^{i\omega})}{(1 + ge^{i\omega} + ge^{2i\omega})} f_y(\omega).$$

---

[23] d_t is a preference shock. We have not discussed the demand side of the model, because we do not use it in estimation.

These equations provide a full description of the model in the frequency domain. σ_ω^2 is a complicated function of the structural parameters, including some from the demand side of the model. All of the parameters of present interest, however, may be identified from the other reduced-form parameters, with the exception of γ_0 and γ_1. We therefore treat σ_ω^2 as a free parameter and estimate it subject to no restrictions.

RMS do not estimate the cattle cycle model. Rather, they choose values for the behavioral parameters and report that the calibrated model fits the data well. In the following section, we explicitly estimate the model and compare our findings to those of RMS.

4c. Assessing, Estimating, and Testing the Model

To assess agreement between a parameterized version of the model and the data, or to estimate parameters formally, it is necessary to construct an explicit loss function. We use a loss function that explicitly incorporates the focus in the cattle cycle literature on cycles of roughly 10 years. The loss function, which measures divergence between model and data spectra only within a particular frequency band, leads us to an estimator that we call band-restricted maximum likelihood (Band-ML). We exclude frequencies corresponding to periods of more than 30 years or less than 4 years.[24] From the standpoint of our earlier discussion of frequency downweighting, this corresponds to weighting frequencies in the band of interest equally, and giving frequencies outside the band zero weight.

In Figure 5 we display the model spectrum evaluated at the Band-ML parameter estimates. Given the objective of constructing a simple model that is consistent with periodic

[24] Gaussian Band-ML is the maximum likelihood analog of Engle's (1974) band-spectral linear regression. Band-ML may of course be undertaken for models much more complicated than simple linear regression, such as the present one.

behavior in these series, a surprising finding is that neither the consumption nor the stock model spectrum has a peak corresponding to a ten-year cycle. Instead, the main distinguishing feature of both model spectra is Granger's (1966) classic spectral shape. This suggests that at the band-ML optimum, the model does not easily produce cyclical behavior. The model phase shift also declines monotonically, which contrasts somewhat with the point estimate of the phase shift, which has a local peak at roughly the ten-year cycle. Finally, the model coherence reminds us of yet another of the model's limitations: because it is driven by a single shock, the model is singular, which produces unit coherence at all frequencies regardless of the parameter configuration.

To evaluate divergence between model and data, we plot the model spectrum in Figure 6, together with the earlier-discussed 90% confidence tunnels for the data spectrum, produced with 200 replications of the non-parametric Cholesky factor bootstrap.[25] The diagonal elements provide comparative assessments of model and data univariate dynamics, and the off-diagonal elements provide comparative assessments of cross-variable dynamics.

Figure 6 reveals some divergence between model and data beyond the earlier-discussed fact that the model spectrum fails to display the internal spectral peaks found in the data spectrum. First, the rate of decay of the model consumption spectrum appears significantly slower than that of the data spectrum; thus, although the consumption model and data spectra agree over most of the relevant frequency range, they begin to deviate substantially for cycles with periods of 4 years or less. Second, and conversely, the rate of

[25] When constructing the bootstrap confidence tunnel, we apply a Bartlett window to the off-diagonal elements of the covariance matrix, and we use a truncation lag of 24.

decay of the model stock spectrum appears significantly *faster* than that of the data spectrum. The two diverge not only at high frequencies, but also over much of the relevant frequency range. In particular, the model stock spectrum lies slightly outside the lower region of the 90% confidence tunnel for cycles of about 20 years and less. Third, the phase shift implied by the model tends to be significantly larger than the phase shift found in the data over the frequencies of interest. Finally, model and data coherence diverge; in spite of the fact that the confidence tunnel is very wide, the unit model coherence is always outside the confidence tunnel for the data coherence.

Let us now discuss the band-ML estimation in greater detail. We estimate model parameters using the simplex algorithm, which is a derivative-free method, as implemented in the Matlab fmins.m procedure. Using penalty functions, we constrain the discount factor to be between 0.65 and 1.00, the fertility rate to be between 0.00 and 1.00, the death rate to be between 0.00 and 1.00, the persistence parameter to be between 0.00 and 1.00, and the scale parameter to be between 0.10 and 7.00. We start the iterations with the RMS parameter values for the discount rate, fertility rate, death rate, and persistence parameter. [26] In our experience, estimation is numerically straightforward and stable; the estimated parameter vector is always in the interior of the constraint set, convergence is fast, and alternative starting values produce virtually identical estimates. In contrast, the RMS model has proven to be quite difficult to estimate using more traditional approaches. For example, Hansen, McGrattan, and Sargent (1997) find that standard time-domain ML fails to converge unless the discount factor is fixed prior to estimation.

---

[26] RMS did not report a value for the scale parameter; we start it at 1.7.

We present the Band-ML estimates and the RMS parameter values in Table 1. We have two main findings. First, several of the parameter values obtained by band-restricted maximum likelihood are similar to those chosen by RMS. In particular, the estimate of the death rate parameter (.08) is nearly identical to the RMS value (0.10), and the estimate of the producer's discount factor (.86) is close to the RMS value (0.91). The estimated fertility parameter (0.67) is lower than but nevertheless close to the RMS value (0.85), which RMS choose based on biological considerations.

Our second main finding is that the band ML estimate of the persistence parameter, which is a fundamental object in the RMS model, differs substantially from the RMS value. RMS choose a fairly persistent value of 0.6. In contrast, we find that optimizing the band-ML loss function requires very little persistence in the driving process (0.2). This implies that the RMS model has a strong internal propagation mechanism: the model takes shocks with relatively little serial correlation and transforms them into series that display substantial persistence in equilibrium. This dimension of the RMS model differs fundamentally from standard dynamic equilibrium models used in macroeconomics, international economics, and public finance. As Watson (1993) and others have noted, models used in those fields typically have weak internal propagation mechanisms -- they require highly persistent underlying shocks to generate a realistic amount of serial correlation in the variables determined in equilibrium. This is considered to be a shortcoming of the models and is the focus of much current research. Thus, a potentially important contribution of the RMS model is that the rich nature of its dynamic propagation mechanisms may be adapted to help researchers in other fields construct models with stronger internal propagation.

In addition to finding the parameter estimates that maximize agreement between model and data, we can assess their sampling uncertainty within our framework. Standard errors are of some use in that regard, in spite of the fact that the sampling distributions need not be Gaussian. We compute them using 200 replications of the Cholesky factor bootstrap procedure, and we report them in parentheses below the estimated parameters in Table 1. More generally, our bootstrap procedures allow us to estimate the *entire* sampling distributions of the estimated parameters; we report in them in Figure 7. The estimated sampling distributions of the discount factor, the depreciation rate, and the persistence parameter are fairly concentrated, while the estimated sampling distribution of fertility rate is more dispersed.

Our framework also enables us to examine the joint distribution of the estimated parameters. In Table 2 we present bootstrap estimates of the correlations between the estimated parameters. Perhaps the most interesting relationship is the strong negative correlation between the discount factor and the fertility rate, which occurs because the discount factor and the fertility rate enter multiplicatively in one of the cubic equations that define the ARMA polynomials. This implies that the loss function trades off high fertility rates for low discount factors, and suggests that fixing either one of the parameters at the higher RMS value would tend to result in an even *lower* estimate for the other.

5. Concluding Remarks

We have described a framework for evaluating dynamic economic models that should be useful to applied economists in many fields. The framework is flexible -- it can be used by researchers to formally evaluate purely calibrated models, and it can also be used by

researchers interested in estimating parameters and conducting inference. Moreover, it is graphical and constructive, and it takes seriously several important issues in the quantitative analysis of simple, dynamic equilibrium models: model misspecification, the user's objectives, and small sample sizes. Its frequency-domain foundations provide useful diagnostics that nicely complement alternative time-domain approaches, such as Canova, Finn and Pagan's (1994) approach based on estimated VARs.

Our analysis of the RMS model of cattle cycles illustrated the use of our tools for assessing agreement between models and data at pre-set parameter values, as well as for formally estimating models and performing statistical inference. In addition, it shed new light on the characteristics of the RMS model, and in particular, its strong internal propagation mechanism. Our analysis also revealed several deficiencies of the model, not the least of which is its inability to generate internal spectral peaks in the model spectra evaluated at the band-ML estimates.

The ultimate goal of the research program of which this paper is a part is to facilitate communication between researchers with potentially very different research objectives and strategies, thereby bringing modern dynamic economic theory into closer and more frequent contact with dynamic economic data. As economists use richer and more complicated models to understand a wider variety of data, we hope that our framework will find use in discerning the dimensions along which models are consistent -- and inconsistent -- with data. That information can in turn be used to construct new and improved models.

References

Auerbach, A.J. and Kotlikoff, L.J. (1987), *Dynamic Fiscal Policy*. Cambridge: Cambridge University Press.

Backus, D.K., Kehoe, P.J. and Kydland, F.E. (1994), "Dynamics of the Trade Balance and the Terms of Trade: The J-Curve ?," *American Economic Review*, 84, 84-103.

Canova, F., Finn, M., and Pagan, A.R. (1994), "Evaluating a Real Business Cycle Model," in C.P. Hargreaves (ed.), *Nonstationary Time Series and Cointegration*. Oxford: Oxford University Press.

Christiano, L.J, and Eichenbaum, M. (1995), "Liquidity Effects, Monetary Policy and the Business Cycle," *Journal of Money, Credit and Banking*, 27, 1113-1136.

Cogley, T. and Nason, J.M. (1995), "Output Dynamics in Real Business Cycle Models," *American Economic Review*, 85, 492-511.

Cooley, T.F. and Hansen, G. (1992), "Tax Distortions in a Neoclassical Monetary Economy," *Journal of Economic Theory*, 58, 290-316.

Engle, R.F. (1974), "Band Spectrum Regression," *International Economic Review*, 15, 1-11.

Ericson, R. and Pakes, A. (1995), Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53-82.

Gallant, A.R., Rossi, P.E. and Tauchen, G. (1993), "Nonlinear Dynamic Structures," *Econometrica*, 61, 871-908.

Granger, C.W.J. (1966), "The Typical Spectral Shape of an Economic Variable," *Econometrica,* 34, 150-161.

Greenwood, J., Hercowitz, Z. and Krusell, P. (1997), "Long Run Implications of Equipment-Specific Technological Change," *American Economic Review*, in press.

Gregory, A.W. and Smith, G.W. (1991), "Calibration as Testing," *Journal of Business and Economic Statistics*, 9, 297-303.

Hannan, E.J. (1970), *Multiple Time Series*. New York: John Wiley.

Hansen, G.D. (1985), "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, 16, 309-327.

Hansen, L.P and Heckman, J.J. (1996), "The Empirical Foundations of Calibration," *Journal*

of Economic Perspectives, 10, 87-104.

Hansen, L.P., McGrattan, E. and Sargent, T.J. (1997), "Mechanics of Forming and Estimating Dynamic Linear Economies," in J. Rust, D. Kendrick, and H. Amman (eds.), *Handbook of Computational Economics*, in press.

Hansen, L.P. and Sargent, T.J. (1993), "Seasonality and Approximation Errors in Rational Expectations Models," *Journal of Econometrics*, 55, 21-55.

Jones, L. and Manuelli, R. (1990), "A Convex Model of Equilibrium Growth: Theory and Policy Implications," *Journal of Political Economy*, 98, 1008-1038.

Kim, K. and Pagan, A.R. (1994), "The Econometric Analysis of Calibrated Macroeconomic Models," in H. Pesaran and M. Wickens (eds.), *Handbook of Applied Econometrics*. Oxford: Blackwell.

King, R.G. and Watson, M.W. (1992), "On the Econometrics of Comparative Dynamics," Manuscript, Departments of Economics, University of Rochester and Northwestern University.

King, R.G. and Watson, M.W. (1996), "Money, Prices, Interest Rates and the Business Cycle," *Review of Economics and Statistics*, 78, 35-53.

Kydland, F. and Prescott, E.C. (1982), "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345-1370.

Kydland, F.E. and Prescott, E.C. (1996), "The Computational Experiment: An Econometric Tool," Journal of Economic Perspectives, 10, 69-86.

Leeper, E.M. and Sims, C.A. (1994), "Toward a Modern Macroeconomic Model Useable for Policy Analysis," in O. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual*, 81-117.

Lucas, R.E. (1977), "Understanding Business Cycles," *Carnegie-Rochester Series on Public Policy*, 5, 7-29.

Lucas, R.E. (1988), "On The Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3-42.

Lucas, R.E. (1990), "Supply Side Economics: An Analytical Review," Manuscript, Department of Economics, University of Chicago.

Manski, C. (1988), *Analog Estimation Methods in Econometrics*. New York: Chapman and Hall.

Ohanian, L.E. (1997), "The Macroeconomic Effects of War Finance in the United States: World War II and the Korean War," *American Economic Review*, in press.

Pagan, A. (1994), "Calibration and Econometric Research: An Overview," in A. Pagan (ed.), *Calibration Techniques and Econometrics*, special issue of *Journal of Applied Econometrics*, 9, S1-S10.

Priestly, M.B. (1981), *Spectral Analysis and Time Series*. London: Academic Press.

Ramos, E. (1988), "Resampling Methods for Time Series," Technical Report ONR-C-2, Department of Statistics, Harvard University.

Rebelo, S. (1991), "Long Run Policy Analysis and Long Run Growth," *Journal of Political Economy*, 99, 500-521.

Rosen, S., Murphy, K.M. and Scheinkman, J.A. (1994), "Cattle Cycles," *Journal of Political Economy*, 102, 468-492.

Rotemberg, J. and Woodford, M. (1996), "Real Business Cycle Models and Forecastable Movements in Output," *American Economic Review*, 86, 71-89.

Rust, J. (1989), "A Dynamic Programming Model of Retirement Behavior," in D. Wise (ed.), *The Economics of Aging*. Chicago: University of Chicago Press.

Sims, C.A. (1993), "Rational Expectations Modeling with Seasonally Adjusted Data," *Journal of Econometrics*, 55, 9-19.

Sims, C.A. (1996), "Macroeconomics and Methodology," *Journal of Economic Perspectives*, 10, 105-120.

Swanepoel, J.W.H. and van Wyk, J.W.J. (1986), "The Bootstrap Applied to Power Spectral Density Function Estimation," *Biometrika*, 73, 135-141.

Watson, M.W. (1993), "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101, 1011-1041.

Woodroofe, M.B. and van Ness, J.W. (1967), "The Maximum Deviation of Sample Spectral Densities," *Annals of Mathematical Statistics*, 38, 1559 -1569.

Table 1
Parameter Estimates
Band-Restricted Maximum Likelihood Estimation

Parameter:	β	g	δ	ρ	σ^2_ω
Estimation or Calibration Method					
Band-ML	.86	.67	.08	.21	2.10
	(.03)	(.09)	(.03)	(.10)	(.37)
RMS	.909	.85	.10	.60	NA
	(NA)	(NA)	(NA)	(NA)	(NA)

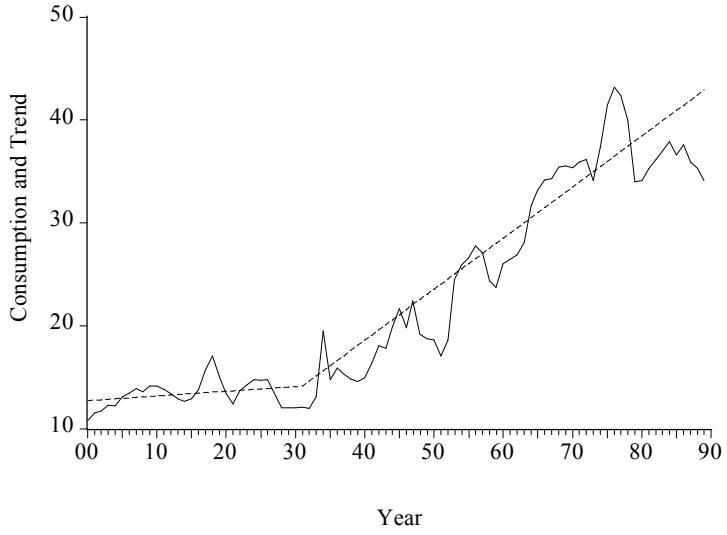
Notes to Table: β is the discount factor, g is the is the fertility rate, δ is the death rate, and ρ is the persistence parameter. Band-ML denotes band-restricted maximum likelihood estimation, with the frequency band used for estimation corresponding to periods from 30 to 4 years. Standard errors, based on 200 bootstrap replications, appear in parentheses. RMS denotes the Rosen-Murphy-Scheinkman calibrated parameters. (They have no standard errors, because they were not estimated.)

Table 2
Estimated Parameter Correlations
Band-Restricted Maximum Likelihood Estimation

	β	g	δ	ρ
β	1.00			
g	-.73	1.00		
δ	.49	-.37	1.00	
ρ	-.19	.10	.06	1.00

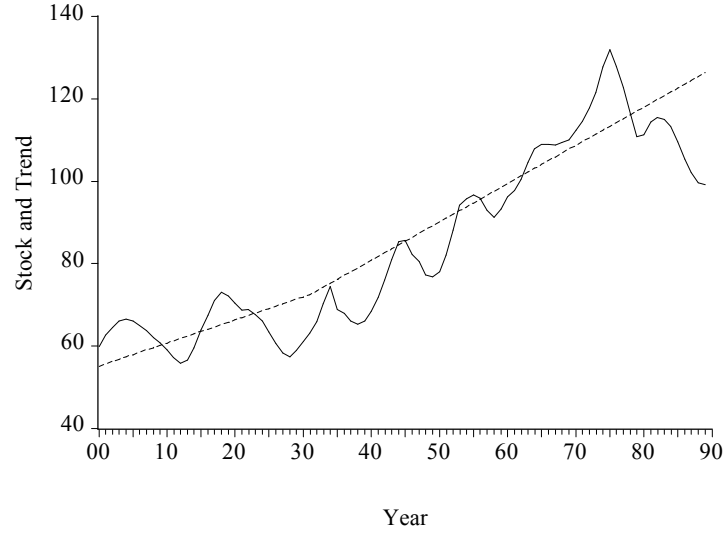
Notes to Table: β is the discount factor, g is the fertility rate, δ is the death rate, and ρ is the persistence parameter. Estimated parameter correlations are based on 200 bootstrap replications. The frequency band used for estimation corresponds to periods from 30 to 4 years.

Figure 1
U.S. Cattle Consumption, 1900-1990
Actual and Estimated Trend



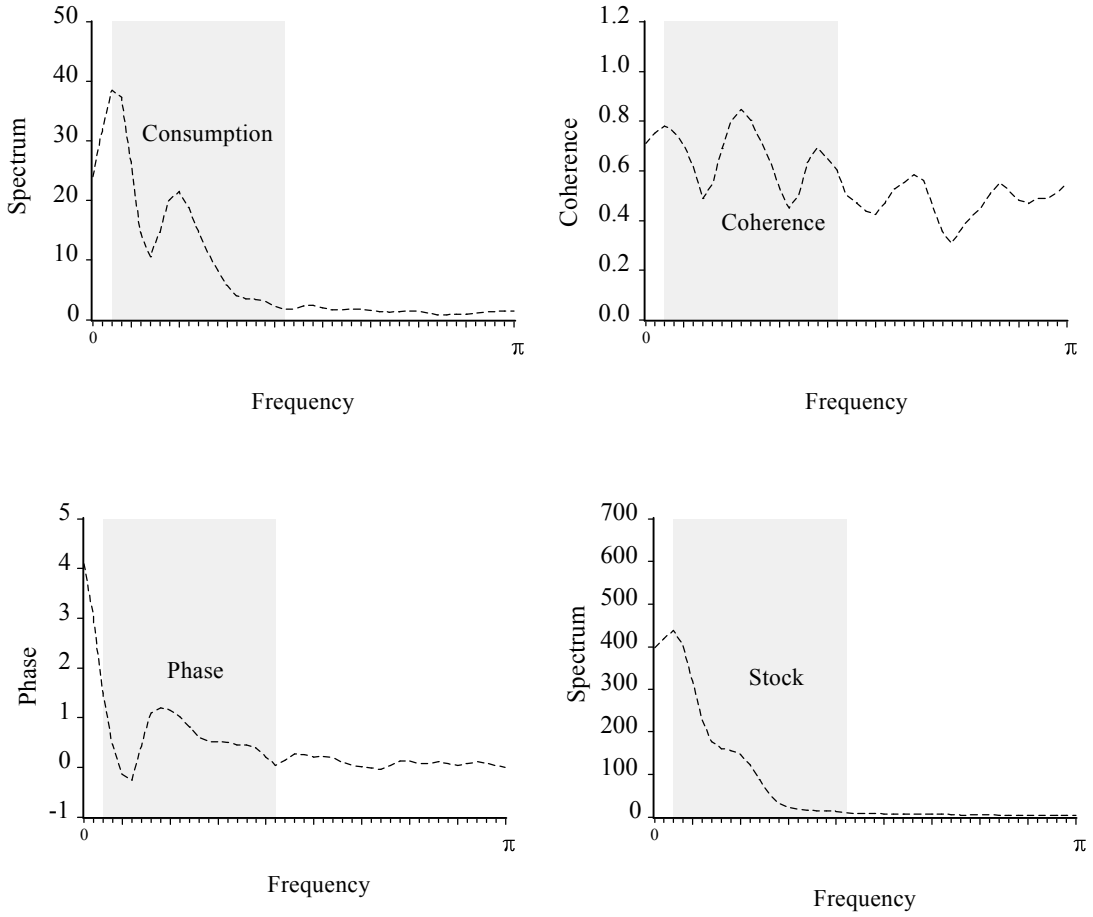
Notes to Figure: We show cattle consumption (solid line) and the estimated kinked-linear trend (dashed line).

Figure 2
U.S. Cattle Stock, 1900-1990
Actual and Estimated Trend



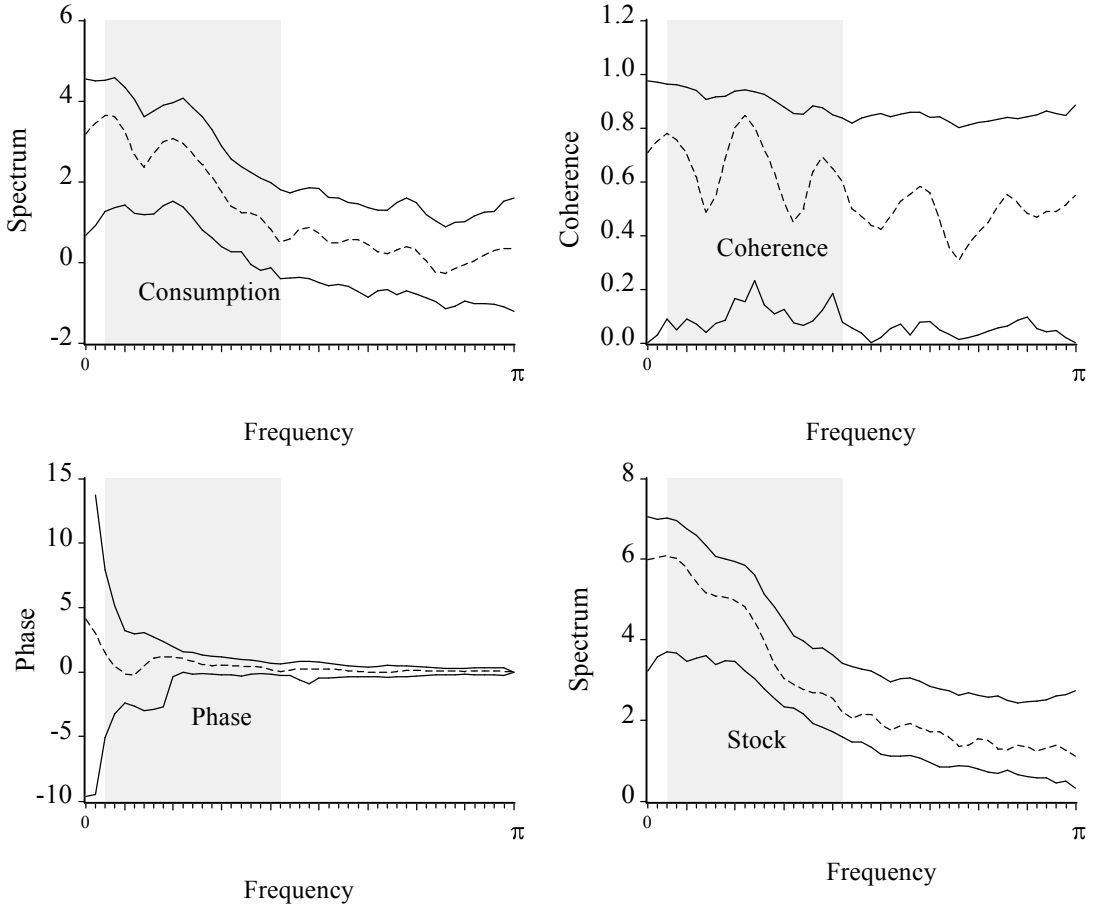
Notes to Figure: We show cattle stock (solid line) and the estimated kinked-linear trend (dashed line).

Figure 3
Estimated Spectral Density Matrix
U.S. Cattle Consumption and Stock



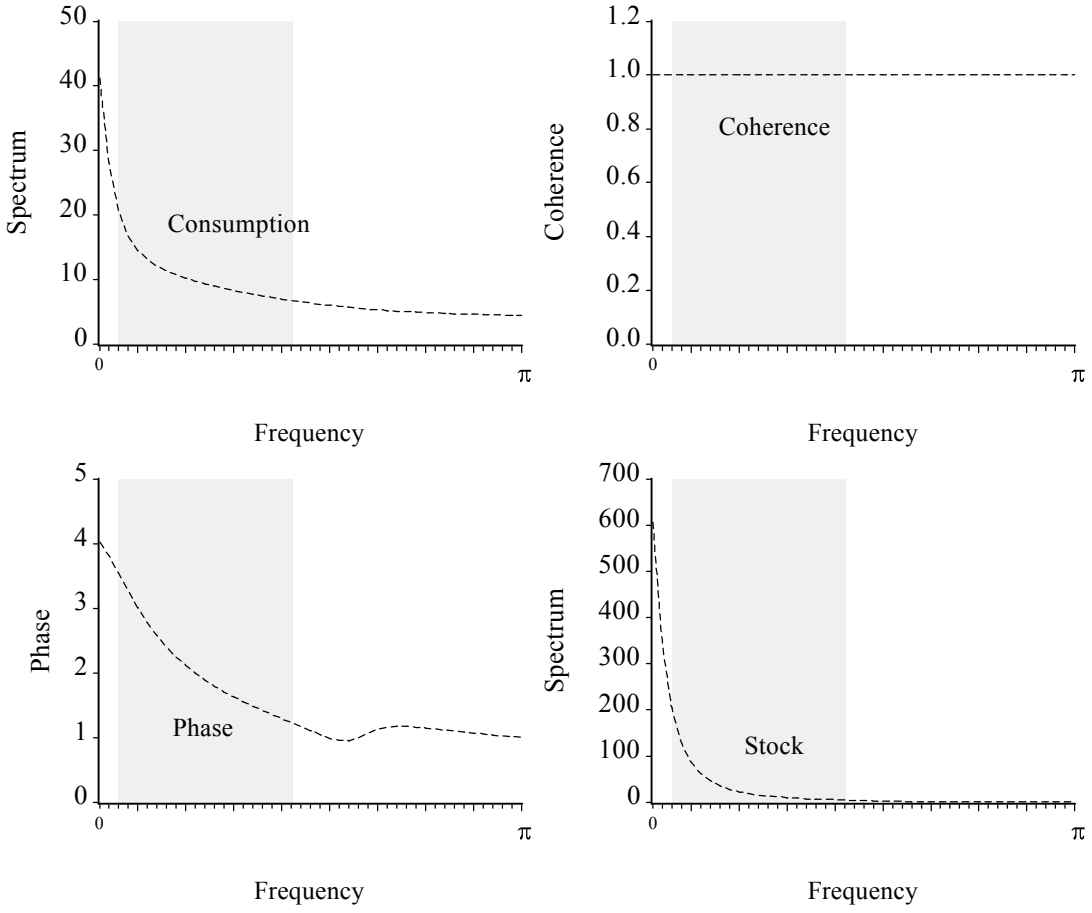
Notes to Figure: We detrend all data using the kinked-linear method. We show the point estimate of each element of the spectral density matrix. The frequency band indicated by vertical dashed lines corresponds to cycles with periods of 30 to 4 years and is the band of primary relevance for studying cattle cycles.

Figure 4
Estimated Spectral Density Matrix and Confidence Tunnels
U.S. Cattle Consumption and Stock



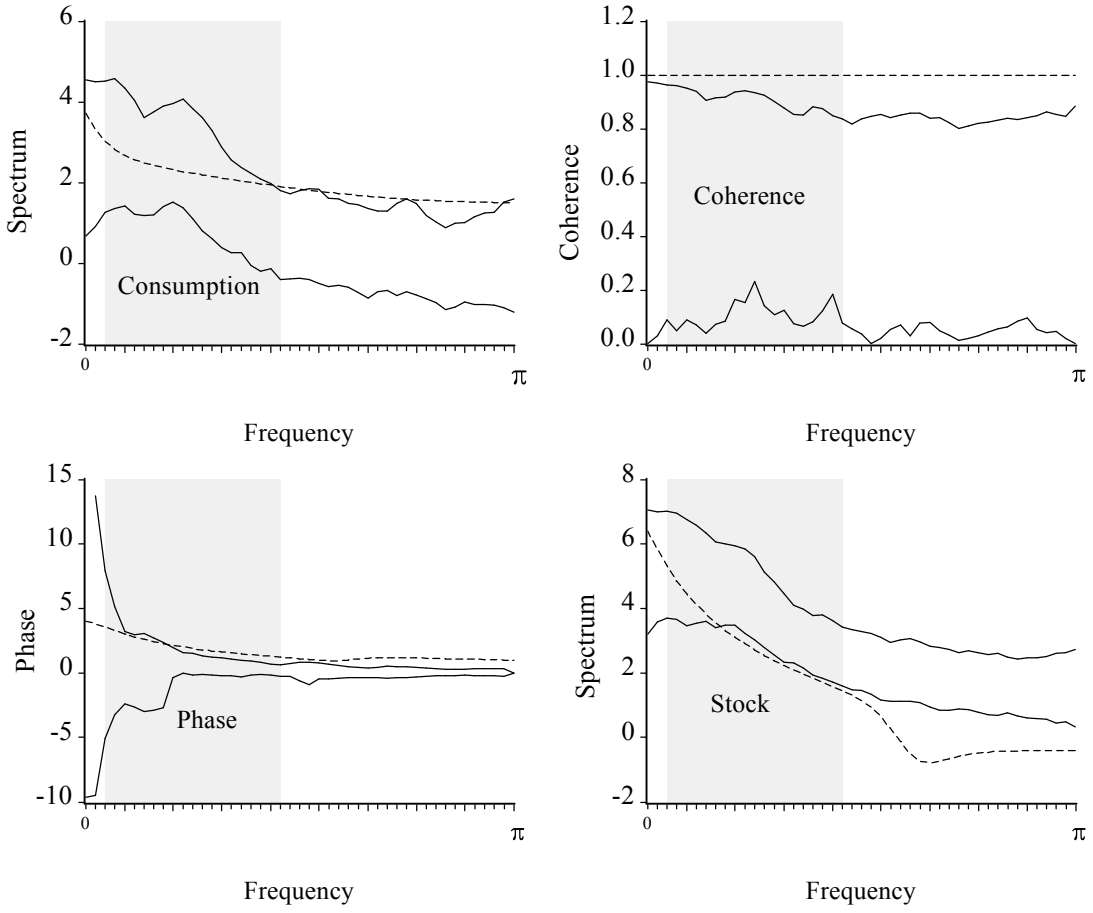
Notes to Figure: We detrend all data using the kinked-linear method. We show the poin t estimate together with a 90% confidence tunnel for each element of the spectral density matrix. The frequency band indicated by vertical dashed lines corresponds to cycles with periods of 30 to 4 years and is the band of primary relevance for studying cattle cycles.

Figure 5
Model Spectrum Evaluated at Band-ML Estimates
U.S. Cattle Consumption and Stock



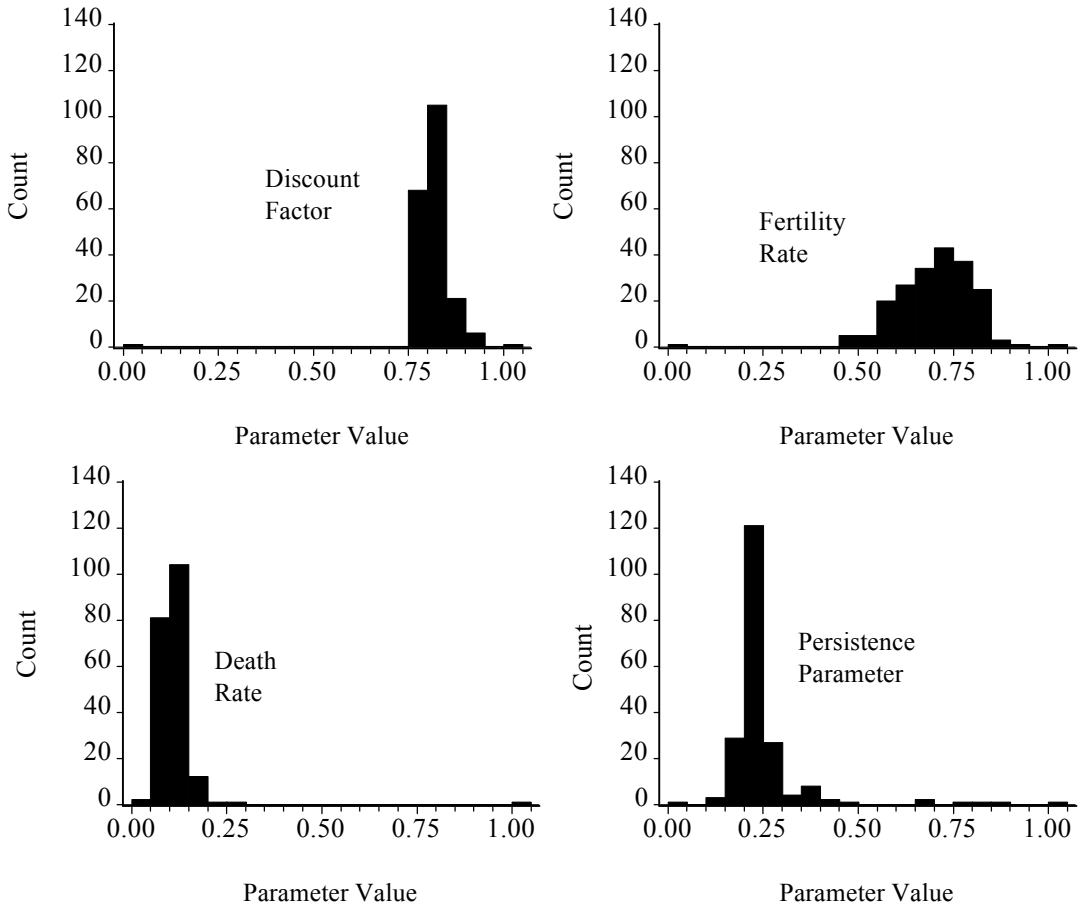
Notes to Figure: We detrend all data using the kinked-linear method. We show the model spectrum evaluated at the band-restricted maximum likelihood parameter values, for each element of the spectral density matrix. The frequency band indicated by vertical dashed lines corresponds to cycles with periods of 30 to 4 years and is the band of primary relevance for studying cattle cycles.

Figure 6
Model Spectra, and Data Spectra Confidence Tunnels
U.S. Cattle Consumption and Stock



Notes to Figure: We detrend all data using the kinked-linear method. We show the 90% confidence tunnel for the data spectrum, together with the model spectrum evaluated at the band-restricted maximum likelihood parameter values, for each element of the spectral density matrix. The frequency band indicated by vertical dashed lines corresponds to cycles with periods of 30 to 4 years and is the band of primary relevance for studying cattle cycles.

Figure 7
Bootstrap Estimates of Sampling Distributions



Notes to Figure: Estimated sampling distributions are based on 200 bootstrap replications.

Appendix 1

Asymptotic Properties of the Cholesky Factor Bootstrap[27]

Under a normality assumption, Ramos (1988) proves first-order validity of what is essentially a frequency-domain variant of the Cholesky factor bootstrap for smooth functionals of the spectrum, but ironically enough, his proof does not cover the spectrum itself, on which our attention centers. In this appendix, we establish first-order asymptotic validity of the Cholesky factor bootstrap for the spectrum. The intuition is simple enough. In the finite-ordered MA(q) case in which a bound on the order $m \geq q$ is known, the Cholesky factor bootstrap is obviously valid, as all autocovariances beyond displacement m are known to be zero, which can be imposed in construction of $\hat{\Sigma}$. In the general case, which via Wold's theorem corresponds to an infinite-ordered moving average with square summable coefficients, the key is to allow the number and impact of the sample autocovariances included in the construction of $\hat{\Sigma}$ to grow with sample size, but at a slower rate. Hence the downweighting associated with the use of Σ^* rather than $\hat{\Sigma}$.[28]

1. Background and Assumptions

Let $\{y_t, t = 1, ..., T\}$ be the sample path of a covariance stationary Gaussian time series. For simplicity, we assume that y_t is a scalar random variable with known mean, which we take as 0 without loss of generality. Let

---

[27] This appendix was written by J. Hahn (Department of Economics, University of Pennsylvania) and F.X. Diebold.

[28] Many downweighting schemes are admissible; the situation is precisely analogous to the variety of admissible windows for estimating spectra.

$$\sigma(r) = E\ [y_t\ y_{t+r}]$$

$$C_r = \frac{1}{T - |r|} \sum y_t\ y_{t+r}$$

$$c_r = \frac{1}{T} \sum y_t\ y_{t+r}$$

$$I(\lambda) = \frac{1}{2\pi T} \left| \sum_t y_t e^{i\lambda t} \right|^2 .$$

We estimate the spectral density nonparametrically in the usual way as

$$\hat{f}(\nu) = \frac{1}{2\pi} \sum_{r\ =\ -(T-1)}^{T-1} w_{rT} \cos(\nu r) C_r = \frac{1}{2\pi} \sum_{r\ =\ -(T-1)}^{T-1} w_{rT}^{*} \cos(\nu r) c_r$$

for

$$w_{rT}^{*} = \frac{T}{T - |r|} w_{rT} .$$

We usually use

$$w_{rT}^{*} = \begin{cases} k\left(\frac{r}{K_T}\right) & \text{if } |r| \le K_T \\ 0 & \text{otherwise,} \end{cases}$$

where the kernel k (.) is a continuous symmetric function with k (0) = 1.

Assumption. We assume that

$$\lim_{x \to 0} \frac{1 - k(x)}{|x|^q} = k$$

for some strictly positive q and k.

Assumption. We assume that

$$\sum_{r=-\infty}^{\infty} |r|^p |\sigma(r)| < \infty$$

for some strictly positive p.

Assumption. We assume that

$$K_T \to \infty$$

and

$$\frac{K_T^m}{T} \to 0,$$

where m = min (p,q).

The asymptotic properties of our spectral density estimator are well known. Anderson (1971), for example, discusses them in detail, and for convenience of exposition we will pattern this appendix closely after Anderson's Chapter 9. In particular, we will characterize the "bootstrap world" distribution using precisely the same flow of logic that Anderson uses to characterize the "real world" distribution.

2. The Bootstrap World

We rely on the triangular nature of the bootstrap, which implies that we need not even consider the bootstrap explicitly. Instead, we need only consider a triangular array of univariate time series that converges to the original time series. We will invoke a number of regularity conditions as we proceed, and we will verify them later.

Condition. $\{y_{tT},\ t=1,...,T\}$ is a triangular array of zero mean stationary Gaussian random variables.

Let

$$\sigma_T(r) = E\left[y_{tT} \cdot y_{t+r,\,T}\right]$$

$$C_{rT} = \frac{1}{T-|r|} \sum y_{tT} \cdot y_{t+r,\,T}$$

$$c_{rT} = \frac{1}{T} \sum y_{tT} \cdot y_{t+r,\,T}$$

$$\hat{f}_T(v) = \frac{1}{2\pi} \sum_{r=-(T-1)}^{T-1} w_{rT} \cos(vr) C_{rT}$$

$$= \frac{1}{2\pi} \sum_{r=-(T-1)}^{T-1} w^*_{rT} \cos(vr) c_{rT}.$$

Condition. $\sigma_T(r) \to \sigma(r)$ for every r, and

$$\lim_{T\to\infty} \sum_{r=-\infty}^{\infty} |r|^p |\sigma_T(r)| = \sum_{r=-\infty}^{\infty} |r|^p |\sigma(r)| < \infty .$$

Condition. For any $K_T^* = O(K_T)$,

$$\lim_{T\to\infty} \sum_{r=K_T^*}^{\infty} |r|^m |\sigma_T(r)| = 0.$$

Lemma 1

$$\lim_{T\to\infty} \hat{f}_T(v) = f(v).$$

The proof of the lemma is trivial and we omit it.

2a. Asymptotic Bias

We first consider the asymptotic bias of $\hat{f}_T(v)$. By Anderson's equation (24), p. 524, we have

$$
\begin{aligned}
K_T^m \left[E\hat{f}_T(v) - f_T(v)\right] &= \frac{K_T^m}{2\pi} \sum_{r=-K_T}^{K_T} \left[k\left(\frac{r}{K_T}\right)\left(1 - \frac{|r|}{T}\right) - 1\right] \cos(vr) \cdot \sigma_T(r) \\
&\quad - \frac{2K_T^m}{2\pi} \sum_{r=K_T+1}^{\infty} \cos(vr) \cdot \sigma_T(r) \\
&= \frac{K_T^m}{2\pi} \sum_{r=-K_T}^{K_T} \left[k\left(\frac{r}{K_T}\right) - 1\right] \cos(vr) \cdot \sigma_T(r) \\
&\quad - \frac{K_T^m}{\pi} \sum_{r=1}^{K_T} k\left(\frac{r}{K_T}\right) \frac{r}{T} \cos(vr) \cdot \sigma_T(r) \\
&\quad - \frac{K_T^m}{\pi} \sum_{r=K_T+1}^{\infty} \cos(vr) \cdot \sigma_T(r).
\end{aligned} \tag{1}
$$

The first term on the right can be written

$$
\begin{aligned}
&\frac{1}{2\pi} \sum_{r=-K_T^*}^{K_T^*} \frac{k\left(\frac{r}{K_T}\right) - 1}{\left|\frac{r}{K_T}\right|^m} |r|^m \cos(vr) \cdot \sigma_T(r) \\
&+ \frac{1}{\pi} \sum_{r=K_T^*+1}^{K_T} \frac{k\left(\frac{r}{K_T}\right) - 1}{\left|\frac{r}{K_T}\right|^m} |r|^m \cos(vr) \cdot \sigma_T(r)
\end{aligned} \tag{2}
$$

for any integer $K_T^* < K_T$. Because

$$
\lim_{x \to 0} \frac{1 - k(x)}{|x|^q} = k,
$$

for any $\epsilon > 0$, we can choose $\delta > 0$ so that

$$
\left| \frac{k\left(\frac{r}{K_T}\right) - 1}{\left|\frac{r}{K_T}\right|^q} + k \right| < \epsilon \text{ if } \left|\frac{r}{K_T}\right| < \delta.
$$

Now set $K_T^* = [\delta K_T]$. If $m = q \leq p$, the first term in (2) is within

$$\epsilon' = \epsilon \cdot \sum_{r=-\infty}^{\infty} \frac{|r|^q |\sigma_T(r)|}{2\pi}$$

of

$$-\frac{k}{2\pi} \sum_{r=-K_T^*}^{K_T^*} |r|^q \cos(\nu r) \cdot \sigma_T(r),$$

which converges to $-kf^{[q]}(\nu)$, where

$$f^{[q]}(\nu) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} |r|^q \cos(\nu r) \cdot \sigma(r).$$

If $m = p < q$, the first term in (2) is within ϵ' of

$$-\frac{k}{2\pi} \sum_{r=-K_T^*}^{K_T^*} \left|\frac{r}{K_T}\right|^{q-p} |r|^p \cos(\nu r) \cdot \sigma_T(r),$$

which in absolute value is no greater than

$$\delta^{q-p} \frac{k}{2\pi} \sum_{r=-\infty}^{\infty} |r|^p |\sigma_T(r)|,$$

and hence arbitrarily small. If $|k(x)| \leq M$, then

$$\frac{|1-k(x)|}{|x|^m} \leq \frac{M+1}{\delta^m}, \qquad |x| \geq \delta.$$

The second term in (2) is in absolute value no greater than

$$\frac{M+1}{\pi\delta^m} \sum_{r=[\delta K_T]+1}^{\infty} r^m |\sigma_T(r)|,$$

which converges to 0. The second term on the right hand side of (1) is in absolute value no greater than

$$M \frac{K_T^m}{\pi T} \sum_{r=1}^{K_T} r|\sigma_T(r)|,$$

which converges to 0 if $K_T^m/T \to 0$. The third term on the right hand side of (1) is an absolute value no greater than

$$\frac{K_T^m}{\pi} \sum_{r=K_T+1}^{\infty} |\sigma_T(r)| \le \frac{1}{\pi} \sum_{r=K_T+1}^{\infty} |r|^p |\sigma_T(r)|,$$

which converges to 0. We thus conclude that the asymptotic bias of the triangular array is the same as that given in Anderson.

2b. Asymptotic Variance

Now we consider the asymptotic variance. Using Anderson's equation (44) we write

$$\frac{T}{K_T} \text{Cov}\,[\hat{f}_T(\lambda), \hat{f}_T(\nu)]$$

$$= \frac{1}{(2\pi)^2 K_T} \sum_{g,h=-K_T}^{K_T} k\left(\frac{g}{K_T}\right) k\left(\frac{h}{K_T}\right) e^{i(\lambda g+\nu h)} \qquad (3)$$

$$\times \sum_{r=-(T-1)}^{T-1} \phi_T(r;g,h)\,[\sigma_T(r)\sigma_T(r+h-g) + \sigma_T(r-g)\sigma_T(r+h)].$$

For the exact value of $\phi_T(\cdot)$, see Anderson. Consider

$$\frac{1}{(2\pi)^2 K_T} \sum_{g,h=-K_T}^{K_T} \sum_{r=-(T-1)}^{T-1} \phi_T(r;g,h) k\left(\frac{g}{K_T}\right) k\left(\frac{h}{K_T}\right) e^{i(\lambda g+\nu h)} \sigma_T(r-g)\ \sigma_T(r+h)$$

$$= \frac{1}{(2\pi)^2 K_T} \sum_{u,v=-(K_T+T-1)}^{K_T+T-1} \sum_{r=\max[u-K_T, v-K_T, -(T-1)]}^{\min[u+K_T, v+K_T, T-1]} \phi_T(r;r-u,v-r) \tag{4}$$

$$\times k\left(\frac{r-u}{K_T}\right) k\left(\frac{v-r}{K_T}\right) e^{i(\nu v-\lambda u)+i(\lambda-\nu)r}\ \sigma_T(u)\ \sigma_T(v).$$

The sum over r is 0 if the stated lower limit is greater than the stated upper limit. The difference between (4) and

$$\frac{1}{(2\pi)^2 K_T} \sum_{u,v=-m}^{m} \sum_{r=\max[u-K_T, v-K_T, -(T-1)]}^{\min[u+K_T, v+K_T, T-1]} \phi_T(r;r-u,v-r)$$

$$\times k\left(\frac{r-u}{K_T}\right) k\left(\frac{v-r}{K_T}\right) e^{i(\nu v-\lambda u)+i(\lambda-\nu)r}\ \sigma_T(u)\ \sigma_T(v) \tag{5}$$

is less in absolute value than

$$\left(2+\frac{1}{K_T}\right) \frac{4}{(2\pi)^2} \sum_{u=-\infty}^{\infty} \sum_{v=m+1}^{\infty} \sup_{-1\le x\le 1} k^2(x) |\sigma_T(u)| |\sigma_T(v)|,$$

which is arbitrarily small if $m \le K_T$ is sufficiently large and if $|k(\cdot)|$ is bounded. If k(x) is continuous on [-1, 1], then for $|u| \le m, |v| \le m$ and K_T sufficiently large, we have

$$\left| k\left(\frac{r-u}{K_T}\right) k\left(\frac{v-r}{K_T}\right) - k^2\left(\frac{r}{K_T}\right) \right| < \epsilon$$

for r such that $-K_T \le r-u \le K_T$, $-K_T \le v-r \le K_T$, and $-K_T \le r \le K_T$. From Anderson's equation (49), p. 529, we have, for $|u| \le m, |v| \le m, |r| \le m+K_T$,

$$\phi_T(r;r-u,v-r) \ge 1 - \frac{5m+3K_T}{T}.$$

The difference between (5) and

$$\frac{1}{(2\pi)^2}\sum_{u,v=-m}^{m}\sum_{r=-K_T}^{K_T}\frac{1}{K_T}\,k^2\left(\frac{r}{K_T}\right)e^{i(\nu v-\lambda u)+i(\lambda-\nu)r}\,\sigma_T(u)\sigma_T(v)$$

$$=\left[\sum_{r=-K_T}^{K_T}\frac{1}{K_T}k^2\left(\frac{r}{K_T}\right)e^{i(\lambda-\nu)r}\right]\left[\frac{1}{2\pi}\sum_{u=-m}^{m}e^{-i\lambda u}\sigma_T(u)\right]\left[\frac{1}{2\pi}\sum_{v=-m}^{m}e^{i\nu v}\sigma_T(v)\right] \tag{6}$$

is arbitrarily small if T is sufficiently large. If $\lambda-\nu=0$ or $\pm2\pi$, the limit of the sum on r as $T\to\infty$ is

$$\lim_{K_T\to\infty}\sum_{r=-K_T}^{K_T}\frac{1}{K_T}\,k^2\left(\frac{r}{K_T}\right)=\int_{-1}^{1}k^2(x)\,dx.$$

For m sufficiently large, the limit of (6) is arbitrarily close to

$$f^2(\nu)\int_{-1}^{1}k^2(x)\,dx.$$

If $0<|\lambda-\nu|<2\pi$, then the limit of (6) equals 0 because[29]

$$\sum_{r=-K_T}^{K_T}\frac{1}{K_T}\,k^2\left(\frac{r}{K_T}\right)e^{i(\lambda-\nu)r^r}\to 0.$$

Next from (3), we consider

---

[29] See Anderson's Problem 9.23.

$$\frac{1}{(2\pi)^2 K_T} \sum_{g,h=-K_T}^{K_T} \sum_{r=-(T-1)}^{T-1} \phi_T(r;g,h) k\left(\frac{g}{K_T}\right) k\left(\frac{h}{K_T}\right) e^{i(\lambda g+\nu h)} \sigma_T(r)\,\sigma_T(r+h-g)$$

$$= \frac{1}{(2\pi)^2 K_T} \sum_{u=-(T-1)}^{T-1} \sum_{v=u-2K_T}^{u+2K_T} \sum_{s=\max(u,v)-K_T}^{\min(u,v)+K_T} \phi_T(u;u-s,v-s) \tag{7}$$

$$\times k\left(\frac{u-s}{K_T}\right) k\left(\frac{v-s}{K_T}\right) e^{i(\lambda u+\nu v)-i(\lambda+\nu)s}\ \sigma_T(u)\ \sigma_T(v).$$

We approximate (7) by

$$\frac{1}{(2\pi)^2 K_T} \sum_{u,v=-m}^{m} \sum_{s=\max(u,v)-K_T}^{\min(u,v)+K_T} \phi_T(u;u-s,v-s)$$

$$\times k\left(\frac{u-s}{K_T}\right) k\left(\frac{v-s}{K_T}\right) e^{i(\lambda u+\nu v)-i(\lambda+\nu)s}\ \sigma_T(u)\ \sigma_T(v),$$

which we in turn approximate by

$$\frac{1}{(2\pi)^2} \sum_{u,v=-m}^{m} \sum_{s=-K_T}^{K_T} \frac{1}{K_T} k^2\left(\frac{s}{K_T}\right) e^{i(\lambda u+\nu v)-i(\lambda+\nu)s}\ \sigma_T(u)\ \sigma_T(v), \tag{8}$$

because $k[(u-s)/K_T]k[(v-s)/K_T]$ can be approximated by $k^2(-s/K_T) = k^2(s/K_T)$ for large K_T.

If $\lambda+\nu=0$ or $\pm 2\pi$, the limit of (8) is

$$f^2(v) \int_{-1}^{1} k^2(x)\,dx.$$

Otherwise, the limit is 0. We thus conclude that the asymptotic variance from the triangular

array is the same as that given in Anderson Theorem 9.3.4.

2c. Asymptotic Normality

We now discuss the asymptotic normality of the spectral estimate for the triangular array.

Assumption. We assume that

$$y_{tT} = \sum_{s=-\infty}^{\infty} \gamma_{sT} v_{t-s},$$

where v_t is a sequence of i.i.d. zero mean normal random variables with variance equal to σ^2, and $\gamma_{sT} \to \gamma_s$ pointwise.

We will find the limiting distribution of the difference between

$$U_T = \sqrt{\frac{T}{K_T}} \sum_{g=1}^{K_T} k\left(\frac{g}{K_T}\right) \cos(\nu g) . \frac{1}{T} \sum_{t=1}^{T-g} y_{tT} \cdot y_{t+g,T}$$

and its expectation. The difference between $\sqrt{\frac{T}{K_T}} \left[\hat{f}_T(\nu) - E\hat{f}_T(\nu)\right]$ and $(U_T - EU_T)/\pi$ equals

$$\frac{k(0)}{2\pi} \frac{1}{\sqrt{TK_T}} \sum_{t=1}^{T} \left[y_{t,T}^2 - \sigma_T(0)\right]. \tag{9}$$

From Anderson's Theorem 8.2.4, we find that the variance of

$$\sum_{t=1}^{T} \left[y_{t,T}^2 - \sigma_T(0)\right]$$

equals

$$2 \sum_{r=-(T-1)}^{T-1} (T - |r|) \sigma_T^2(r).$$

It thus follows that the variance of (9) equals

$$\frac{k(0)^2}{(2\pi)^2}\,\frac{1}{K_T}\sum_{r=-(T-1)}^{T-1}\left(1-\frac{|r|}{T}\right)\,\sigma_T^2(r),$$

which converges to 0 as $T\to\infty$.

Now, let

$$y_{tT}^{n}=\sum_{s=-n}^{n}\gamma_{sT}v_{t-s}$$

$$u_{tT}^{n}=\sum_{|s|>n}\gamma_{sT}v_{t-s}$$

$$U_T^{n}=\sqrt{\frac{T}{K_T}}\sum_{g=1}^{K_T}k\left(\frac{g}{K_T}\right)\cos(\nu g)\cdot\frac{1}{T}\sum_{t=1}^{T-g}y_{tT}^{n}y_{t+g,T}^{n}.$$

Then,

$$U_T-U_T^{n}=S_1+S_2+S_3$$

where

$$S_1=\sqrt{\frac{1}{TK_T}}\sum_{g=1}^{K_T}k\left(\frac{g}{K_T}\right)\cos(\nu g)\cdot\sum_{t=1}^{T-g}u_{tT}^{n}y_{t+g,T}^{n}$$

$$S_2=\sqrt{\frac{1}{TK_T}}\sum_{g=1}^{K_T}k\left(\frac{g}{K_T}\right)\cos(\nu g)\cdot\sum_{t=1}^{T-g}y_{tT}^{n}u_{t+g,T}^{n}$$

$$S_3=\sqrt{\frac{1}{TK_T}}\sum_{g=1}^{K_T}k\left(\frac{g}{K_T}\right)\cos(\nu g)\cdot\sum_{t=1}^{T-g}u_{tT}^{n}u_{t+g,T}^{n}.$$

Using Anderson's equations (11) and (12), pp. 535-536, we find that the variance of S_3 is bounded by

$$\sup_{-1\le x\le 1} k^2(x)\cdot 2\sigma^4\cdot\left(\sum_{|s|>n}|\gamma_{sT}|\right)^4.$$

Using Anderson's equation (13), p. 536, we find that the variance of S_1 and S_2 are bounded by

$$\sup_{-1\le x\le 1} k^2(x)\cdot 2\sigma^4\cdot\left(\sum_{|s|\le n}|\gamma_{sT}|\right)^2\left(\sum_{|s|>n}|\gamma_{sT}|\right)^2.$$

Condition. $\Sigma_s\ |\gamma_{sT}|\to\Sigma_s\ |\gamma_s|$ as $T\to\infty$.

Condition. $\Sigma_{|s|>n}\ |\gamma_{sT}|\to 0$ as $n\to\infty$ for every T.

Because the variance of S_1, S_2, S_3 disappears as $n\to\infty$, the limit distribution of U^n_T is the limit of $\lim_{n\to\infty}\lim_{T\to\infty}\mathcal{L}(U^n_T)$. Notice, as in Anderson's equation (15), that U^n_T is the real part of

$$\begin{aligned}&\frac{1}{\sqrt{TK_T}}\sum_{g=1}^{K_T} k\left(\frac{g}{K_T}\right) e^{ivg}\sum_{t=1}^{T-g} y^n_{t,T}y^n_{t+g,T}\\ =&\frac{1}{\sqrt{TK_T}}\sum_{r,s=-n}^{n}\gamma_{rT}e^{ivr}\,\gamma_{sT}e^{-ivs}\sum_{h=s-r+1}^{K_T+s-r} k\left(\frac{h+r-s}{K_T}\right) e^{ivh}\sum_{q=1-s}^{T-h-r} v_q v_{q+h}.\end{aligned}\tag{10}$$

The difference between the real parts of (10) and

$$\frac{1}{\sqrt{TK_T}}\sum_{r,s=-n}^{n}\gamma_{rT}e^{ivr}\,\gamma_{sT}e^{-ivs}\sum_{h=1}^{K_T-2n} k\left(\frac{h+r-s}{K_T}\right) e^{ivh}\sum_{q=1-s}^{T-h} v_q v_{q+h}\tag{11}$$

has a mean square error that goes to 0 as $T\to\infty$, because for given r and s the difference between the summands in (10) and (11) consists of the terms in the sums on h and q that are included in one expression and not in the other. The number of such terms is less than $AK_Tn+BTn+Cn^2$ for suitable A, B, C, the terms are uncorrelated, and the expected value of the square of the real part of each terms is at most

$$\left(\max_{-n \le r \le n} |\gamma_{rT}| \right)^4 \frac{\sup_{-1 \le x \le 1} k^2(x)\sigma^4}{TK_T} \le \left(\sum_{r=-\infty}^{\infty} |\gamma_{rT}| \right)^4 \frac{\sup_{-1 \le x \le 1} k^2(x)\sigma^4}{TK_T}.$$

Hence the expected value of the square of the difference for each r and s goes to zero as $T \to \infty$. If k(x) is continuous on [-1,1], then $k[(h+r-s)/K_T]$ is arbitrarily close to $k(h/K_T)$ for K_T sufficiently large and $|r| \le n$, $|s| \le n$, $|h| \le K_T$ and $|h + r - s| \le K_T$. Thus the difference between the real parts of (11) and

$$\begin{aligned} &\frac{1}{\sqrt{TK_T}} f_{nT}(\nu) \sum_{h=1}^{K_T-2n} k\left(\frac{h}{K_T}\right) e^{i\nu h} \sum_{q=1}^{T-h} v_q v_{q+h} \\ &= \frac{f_{nT}(\nu)}{\sqrt{TK_T}} \sum_{q=1}^{T-1} \sum_{h=1}^{\min(K_T-2n,\,T-q)} k\left(\frac{h}{K_T}\right) e^{i\nu h} v_q v_{q+h}, \end{aligned} \tag{12}$$

where

$$f_{nT}(\nu) = \left| \sum_{r=-n}^{n} \gamma_{rT} e^{i\nu r} \right|^2,$$

has a mean square error that is arbitrarily small. The difference between the real part of (12) and

$$\frac{f_{nT}(\nu)}{\sqrt{T}} \sum_{q=1}^{T} \frac{1}{\sqrt{K_T}} \sum_{h=1}^{K_T} k\left(\frac{h}{K_T}\right) \cos(\nu h) v_q v_{q+h} = f_{nT}(\nu) \sum_{q=1}^{T} W_{qT}, \tag{13}$$

where

$$W_{qT} = \frac{1}{\sqrt{K_T}} \sum_{h=1}^{K_T} k\left(\frac{h}{K_T}\right) \cos(\nu h) v_q v_{q+h},$$

has a mean square error that goes to 0 as T increases. The process W_{qT} is stationary and finitely dependent, with

$$EW_{qT} = 0$$

$$EW_{qT}^2 = \frac{\sigma^4}{2} \sum_{h=1}^{K_T} k^2\left(\frac{h}{K_T}\right)(1+\cos 2\nu h)/K_T \to \frac{\sigma^4}{2}\int_0^1 k^2(x)\,dx, \quad \nu \neq 0, \pm\pi \qquad (14)$$

$$EW_{qT}\ W_{q+r,T} = 0.$$

Hence the variance of (13) is $f_{nT}^2(\nu)$ times (14).

Now let N_T be a sequence of integers such that $K_T / N_T \to 0$ and $N_T / T \to 0$, let M_T be the largest integer in T / N_T, and let

$$Z_{jT} = \frac{1}{\sqrt{N_T}}\left[W_{(j-1)N_T+1,T} + \ldots + W_{jN_T-K_T,T}\right].$$

Then Z_{jT} $j = 1, \ldots, M_T$ are i.i.d. with means zero and variance given by $1 - K_T / N_T$ times (14). Anderson (p. 539) notes that the difference between

$$\frac{1}{\sqrt{M_T}} \sum_{j=1}^{M_T} Z_{jT} \text{ and } \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_{tT}$$

is stochastically negligible as $T \to \infty$ and that the former has a limiting normal distribution. Notice that neither W_{tT} nor Z_{jT} depends on $\{\gamma_{sT}\}$. We thus find that we obtain the desired asymptotic normal distribution of $\sqrt{T/K_T}[\hat{f}_T(\nu) - E\hat{f}_T(\nu)]$.

3. Verifying the Conditions

Let us recall the conditions needed for asymptotic validity of the Cholesky factor bootstrap, and then verify the conditions.

Condition 1. $\{y_{tT}, t=1, \ldots, T\}$ is a triangular array of zero mean stationary Gaussian random variables.

Condition 2a. $\sigma_T(r) \to \sigma(r)$ for every r.

Condition 2b. $\lim_{T\to\infty} \sum_{r=-\infty}^{\infty} |r|^p |\sigma_T(r)| = \sum_{r=-\infty}^{\infty} |r|^p |\sigma(r)|$.

Condition 3. For any $K^*_T = O(K_T)$, $\lim_{T\to\infty} \sum_{r=K_T^*}^{\infty} |r|^m |\sigma_T(r)| = 0$.

Condition 4. $\Sigma_s |\gamma_{sT}| \to \Sigma_s |\gamma_s|$ as $T \to \infty$.

Condition 5. $\Sigma_{|s|>n} |\gamma_{sT}| \to 0$ as $n \to \infty$ for every T.

Conditions 1 and 2a are obviously satisfied, as is Condition 3 so long as $\ell = o(K_T)$, where for the Cholesky factor bootstrap we use $\sigma_T(r) = 1(|r| \le \ell_T) \frac{1}{T} \sum y_t y_{t+r}$, where ℓ_T is an increasing sequence of integers such that $\ell_T \to \infty$ and $\ell_T = o(T)$. To check Condition 4, note that

$$\sum_s |\gamma_{sT}| \le \sqrt{\sum_s |\gamma_{sT}|^2} = \sqrt{\mathrm{Var}(y_{tT})},$$

where the extreme right side of the equation converges to $\sqrt{\mathrm{Var}(y_t)}$. Thus, by the Dominated Convergence Theorem, the asserted convergence holds. To check Condition 5, all we need to notice is that $\{y_{tT}\}$ is a finite moving average process.

Condition 2b is significantly more challenging to verify. It suffices to show that

$$\sum_{r=1}^{\ell} |r|^p \left(\frac{1}{T} \left| \sum y_t y_{t+r} \right| - |\sigma(r)| \right) \to 0.$$

We first show that

$$\sum_{r=1}^{\ell} |r|^p \left(\frac{1}{T} \left| \sum y_t y_{t+r} \right| - \left| \frac{T-r}{T} \sigma(r) \right| \right) \to 0. \tag{15}$$

Observe that

$$\Pr\left[\left|\sum_{r=1}^{\ell} |r|^p \left(\left|\frac{1}{T}\sum y_t y_{t+r}\right| - \left|\frac{T-r}{T}\sigma(r)\right|\right)\right| > \epsilon\right]$$

$$\leq \sum_{r=1}^{\ell} \Pr\left[|r|^p \left(\left\|\frac{1}{T}\sum y_t y_{t+r} - \left|\frac{T-r}{T}\sigma(r)\right|\right\|\right) > \frac{\epsilon}{\ell}\right]$$

$$\leq \sum_{r=1}^{\ell} \Pr\left[\left(\left|\frac{1}{T}\sum y_t y_{t+r} - \frac{T-r}{T}\sigma(r)\right|\right) > \frac{\epsilon}{\ell |r|^p}\right]$$

$$= \sum_{r=1}^{\ell} \Pr\left[\left(\left|\sum (y_t y_{t+r} - \sigma(r))\right|\right) > \frac{T\epsilon}{\ell |r|^p}\right]$$

$$\leq \sum_{r=1}^{\ell} E\left[\left(\sum (y_t y_{t+r} - \sigma(r))\right)^2\right] \frac{\ell^2 |r|^{2p}}{T^2 \epsilon^2}.$$

If $\{y_t\}$ is a mixing sequence with either ϕ (m) of size 2 or α (m) of size $(2+2\eta) / \eta$, $\eta > 0$, and if $E\left[|y_t y_{t+r}|^{2+2\eta}\right] \leq \Delta < \infty$ for all r, then by White (1984), Lemma 6.19, we have

$$E\left[\left(\sum (y_t y_{t+r} - \sigma(r))\right)^2\right] \leq (T-r)\ \Delta^* \leq T\Delta^*$$

for some Δ^* which does not depend on r. It therefore follows that

$$P_r\left[\left|\sum_{r=1}^{\ell} |r|^p \left(\left|\frac{1}{T}\sum y_t y_{t+r}\right| - \left|\frac{T-r}{T}\sigma(r)\right|\right)\right| > \epsilon\right]$$

$$\leq \sum_{r=1}^{\ell} \frac{\Delta^* \ell^2 |r|^{2p}}{T\epsilon^2}$$

$$\leq \frac{\Delta^* \ell^{2p+3}}{T\epsilon^2}.$$

Therefore, as long as $\ell = o\left(T^{\frac{1}{2p+3}}\right)$, (15) converges to 0. Now, because

$$\sum_{r=1}^{\ell} |r|^{p+1}|\sigma(r)| \leq \frac{\ell^p}{T} \sum_{r=1}^{\ell} |r|^p|\sigma(r)| \to 0,$$

we easily obtain

$$\frac{1}{T} \sum_{r=1}^{\ell} |r|^{p+1}|\sigma(r)| \to 0.$$

4. Discussion

We have proved first-order asymptotic validity for the Cholesky factor bootstrap of the spectral density function. Note that we bootstrap the spectral density function directly; in particular, the object bootstrapped is not asymptotically pivotal. Second-order asymptotic refinements are sometimes available when bootstrapping an asymptotically pivotal statistic, as stressed in Hall (1992). The issue of whether or not one should focus on asymptotically pivotal statistics, however, is by no means uncontroversial. Edgeworth expansions, although providing asymptotic refinements, can and sometimes do make things *worse* in small samples, as stressed in Efron and Tibshirani (1993), who generally prefer to bootstrap non-pivotal statistics.

In closing, we mention that the Cholesky factor bootstrap, which has a nonparametric flavor, and alternatives such as the VAR bootstrap, which has a parametric flavor, are in fact closely related. A modern and unifying view, currently the focus of intense research in mathematical statistics, is to interpret various time series bootstraps as sieves (in the sense of Grenander, 1981) whose complexity increases with sample size at a suitable rate. [30] The Cholesky factor bootstrap has a sieve interpretation; the sieve is a spectrum estimated by smoothing an increasing number of sample autocovariances. Some alternative bootstraps

[30] See, for example, Bühlmann (1997) and Bickel and Bühlmann (1996).

such as those based on VARs also have a sieve interpretation; the sieve is an estimated autoregression of increasing length. Thus, asymptotically in T, both the Cholesky factor and VAR bootstraps can be effective algorithms for generating data with the same second-order properties as an observed sample path. Neither is in general "superior" to the other, and both are the subject of ongoing research, as is the "block" bootstrap of Kunsch (1989) and Liu and Singh (1992) as modified for spectra by Politis and Romano (1992), as well as the spectral bootstrap of Franke and Härdle (1992).

References

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*. New York: John Wiley.

Bickel, P. and Bühlmann, P. (1996), "Mixing Property and Functional Central Limit Theorems for a Sieve Bootstrap in Time Series," Manuscript, Department of Statistics, University of California, Berkeley.

Bühlmann, P. (1997), "Sieve Bootstraps for Time Series," *Bernoulli*, in press.

Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Franke, J. and Härdle, W. (1992), "On Bootstrapping Kernel Spectral Estimates," *Annals of Statistics*, 20, 121-145.

Grenander, U. (1981), *Abstract Inference*. New York: John Wiley.

Hall, P (1992), *The Bootstrap and Edgeworth Expansion*. New York: Springer Verlag.

Künsch, H.R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217-1241.

Liu, R.Y. and Singh, K. (1992), "Moving Blocks Jacknife and Bootstrap Capture Weak Dependence," in R. LePage and L. Billard (eds.), *Exploring the Limits of the Bootstrap*. New York: John Wiley.

Politis D.N. and Romano J.P. (1992), "A General Resampling Scheme for Triangular Arrays of α-Mixing Random Variables with an Application to the Problem of Spectral Density Estimation," *Annals of Statistics*, 20, 1985-2007.

Ramos, E. (1988), "Resampling Methods for Time Series," Technical Report ONR-C-2, Department of Statistics, Harvard University.

White, H. (1984), *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

Appendix 2

Finite-Sample Properties of the Cholesky Factor Bootstrap

In this appendix, we describe the results of a Monte Carlo comparison of the finite-sample properties of the Cholesky factor bootstrap and conventional asymptotics. The experiment is small by necessity, as Monte Carlo evaluation of bootstrap procedures is extremely burdensome computationally, but we believe that it sheds some interesting light on the finite-sample performance of the bootstrap.

We use a data-generating process with realistic dynamics, given by

$$y_t = 1.335y_{t-1} - .401y_{t-2} + \varepsilon_t,\ T = 1, \ldots, 100,$$

which corresponds to Rudebusch's (1993) estimate for detrended log GNP and is representative of the dynamics of a typical detrended macroeconomic series.

We examine the empirical coverage of the nominal 80% and 90% intervals constructed using the Cholesky factor bootstrap and conventional asymptotics. We examine two bootstrap intervals, parametric (Gaussian) and nonparametric. At each of 1000 Monte Carlo replications, we apply the Cholesky factor bootstrap with 2000 bootstrap replications. At each bootstrap replication we estimate the spectral density at frequencies $\pi/6$ and $\pi/2$.

In Table A1, we present the empirical coverage rates for bootstrap and asymptotic confidence intervals for three innovation distributions. First, we set $\varepsilon_t \sim$ iid N(0,1). At frequency $\pi/6$, the actual coverage of all three intervals exceeds nominal coverage. However, both the parametric and nonparametric bootstrap coverage rates are much closer to nominal coverage than those of the asymptotic approximation. At frequency $\pi/2$, the asymptotic intervals similarly deliver excessively high coverage rates but the parametric bootstrap interval in particular (and to a lesser extent the nonparametric) display nearly exact

coverage.

Second, we set ε_t to a conditionally Gaussian GARCH(1,1). As expected, the nonparametric bootstrap outperforms the parametric bootstrap in this case. However, neither the nonparametric bootstrap nor the asymptotic approximation appear definitively best in terms of actual coverage.

Finally, the innovation is iid $\chi^2(2)$, normalized to have zero mean and unit variance. As with iid N(0,1) innovations, we find that the asymptotic approximation tends to give rise to excessively wide confidence intervals. At a nominal coverage level of 90%, both bootstraps deliver more accurate coverage rates. At the nominal 80% level, only the parametric bootstrap dominates the asymptotic interval.

Table A1
Empirical Coverage
Bootstrap and Asymptotic Confidence Intervals

	Nominal Coverage	Parametric Bootstrap Interval	Nonparametric Bootstrap Interval	Asymptotic Interval
Gaussian Innovations				
$f(\pi/6)$	.80	.827	.831	.912
	.90	.913	.910	.974
$f(\pi/2)$	.80	.795	.780	.827
	.90	.904	.901	.980
Conditionally Gaussian GARCH(1,1) Innovations				
$f(\pi/6)$	.80	.696	.718	.767
	.90	.808	.838	.845
$f(\pi/2)$	.80	.770	.818	.789
	.90	.863	.905	.924
Standardized Chi-Square Innovations				
$f(\pi/6)$	.80	.843	.862	.913
	.90	.916	.933	.963
$f(\pi/2)$	.80	.798	.852	.824
	.90	.901	.939	.979

Notes to Table: For each innovation distribution, we generate data from an AR(2) with parameters 1.335 and -.401, with sample size T=100. We perform 2000 bootstrap iterations in each of 1000 Monte Carlo trials.

References

Rudebusch, G.D. (1993), "The Uncertain Unit Root in Real GNP," *American Economic Review*, 83, 264-272.