Testing Models of Low-Frequency Variability*

Ulrich K. Müller
Princeton University
Department of Economics
Princeton, NJ, 08544

Mark W. Watson
Princeton University
Department of Economics
and Woodrow Wilson School
Princeton, NJ, 08544

This version: November 2007
First draft: March 2006

Abstract

We develop a framework to assess how successfully standard time series models explain low-frequency variability of a data series. The low-frequency information is extracted by computing a finite number of weighted averages of the original data, where the weights are low-frequency trigonometric series. The properties of these weighted averages are then compared to the asymptotic implications of a number of common time series models. We apply the framework to twenty U.S. macroeconomic and financial time series using frequencies lower than the business cycle.

JEL classification: C22, E32

Keywords: Long Memory, Local-to-Unity, Unit Root Test, Stationarity Test, Business Cycle Frequency, Heteroskedasticity

*The first draft of this paper was written for the Federal Reserve Bank of Atlanta conference in honor of the twenty-fifth anniversary of the publication of Beveridge and Nelson (1981), and we thank the conference participants for their comments. We also thank Tim Bollerslev, David Dickey, John Geweke, Barbara Rossi, two referees and the Editor for useful comments and discussions, and Rafael Dix Carneiro for excellent research assistance. Support was provided by the National Science Foundation through grants SES-0518036 and SES-0617811. Data and replication files for this research can be found at http://www.princeton.edu/~mwatson.

1 Introduction

Persistence and low-frequency variability has been an important and ongoing empirical issue in macroeconomics and finance. Nelson and Plosser (1982) sparked the debate in macroeconomics by arguing that many macroeconomic aggregates follow unit root autoregressions. Beveridge and Nelson (1981) used the logic of the unit root model to extract stochastic trends from macro series, and showed that variations in these stochastic trends were a large, sometimes dominant, source of variability in the series. Meese and Rogoff's (1983) finding that random walk forecasts of exchange rates dominated other forecasts focused attention on the unit root model in international finance. And in finance, interest in the random walk model arose naturally because of its relation to the efficient markets hypothesis (Fama (1970)).

This empirical interest led to the development of econometric methods for testing the unit root hypothesis, and for estimation and inference in systems that contain integrated series. More recently, the focus has shifted towards more general models of persistence, such as the fractional (or long memory) model and the local-to-unity autoregression, which nest the unit root model as a special case, or in the local level model which allows an alternative nesting of the $I(0)$ and $I(1)$ models. While these models are designed to explain low-frequency behavior of time series, fully parametric versions of the models have implications for higher frequency variation, and efficient statistical procedures thus exploit both low and high frequency variations for inference. This raises the natural concern about the robustness of such inference to alternative formulations of higher frequency variability. These concerns have been addressed by, for example, constructing unit root tests using autoregressive models that are augmented with additional lags as in Said and Dickey (1984), or by using various nonparametric estimators for long-run covariance matrices and (as in Geweke and Porter-Hudak (1983) (GPH)) for the fractional parameter. As useful as these approaches are, there still remains a question of how successful these various methods are in controlling for unknown or misspecified high frequency variability.

This paper takes a different approach. It *begins* by specifying the low-frequency band of interest. For example, the empirical analysis presented in Section 4 focuses mostly on frequencies lower than the business cycle, that is periods greater than eight years. Using this frequency cut-off, the analysis then extracts the low-frequency component of the series of interest by computing weighted averages of the data, where the weights are low-frequency trigonometric series. Inference about the low-frequency variability of the series is exclusively based on the properties of these weighted averages, disregarding other aspects of the original data. The number of weighted averages, say q, that capture the low-frequency variability is small in typical applications. For example, only $q = 13$ weighted averages almost completely capture the lower than business cycle variability in postwar macroeconomic time series (for any sampling frequency). This suggests basing inference on asymptotic approximations in which q is fixed as the sample size tends to infinity. Such asymptotics yield a q-dimensional multivariate Gaussian limiting distribution for the weighted averages, with a covariance matrix that depends on the specific model of low-frequency variability. Inference about alternative models or model parameters can thus draw on the well-developed statistical theory concerning multivariate normal distributions.

An alternative to the methods proposed here is to use time domain filters, such as band-pass or other moving average filters, to isolate the low-frequency variability of the data. The advantage of the transformations that we employ is that they conveniently discretize the low-frequency information of the original data into q data points, and they are applicable beyond the $I(0)$ models typically analyzed with moving average linear filters.

There are several advantages to focusing exclusively on the low-frequency variability components of the data. The foremost advantage is that many empirical questions are naturally formulated in terms of low-frequency variability. For example, the classic Nelson and Plosser (1982) paper asks whether macroeconomic series such as real GNP tend to revert to a deterministic trend over periods longer than the business cycle, and macroeconomic questions involving balanced growth involve the covariability of series over frequencies lower than the business cycle. Questions of potential mean-reversion in asset prices or real exchange rates are often phrased in terms of long "horizons" or low frequencies. Because the statistical models studied here were developed to answer these kinds of low-frequency questions, it is natural to evaluate the models on these terms.

In addition, large literatures have developed econometric methods in the local-to-unity framework, and also in the fractional framework. These methods presumably provide reliable guidance for empirical analysis only if, at a minimum, their assumed framework accurately describes the low-frequency behavior of the time series under study. The tests developed here may thus also be used as specification tests for the appropriateness of these methods. Other advantages, including robustness to high frequency misspecification and statistical convenience (because weighted averages are approximately multivariate normal), have already been mentioned.

An important caveat is that reliance on low-frequency methods will result in a loss of information and efficiency for empirical questions involving all frequencies. Thus, for example, questions about balanced growth are arguably properly answered by the approach developed here, while questions about martingale difference behavior involve a constant spectrum over all frequencies, and focusing only on low frequencies entails a loss of information. Said differently, because only $q = 13$ weighted averages of the data are required to effectively summarize the below-business-cycle variability of post-war economic time series, there are obvious limits to what can be learned from a post-war series about low-frequency variability. Thus, for example, in this 13-observation context one cannot plausibly implement a nonparametric study of low-frequency variability. That said, as the empirical analysis in Section 4 shows, much can be learned from 13 observations about whether the data are consistent with particular low-frequency models.

Several papers have addressed other empirical and theoretical questions in similar frameworks. Bierens (1997) derives estimation and inference procedures for cointegration relationships based on a finite number of weighted averages of the original data, with a joint Gaussian limiting distribution. Phillips (2006) pursues a similar approach with an infinite number of weighted averages. Phillips (1998) provides a theoretical analysis of 'spurious regressions' of various persistent time series on a finite (and also infinite) number of deterministic regressors. Müller (2007b) finds that long-run variance estimators based on a finite number of trigonometrically weighted averages is optimal in a certain sense. All these approaches exploit the known asymptotic properties of weighted averages for a given model of low-frequency variability. In contrast, the focus of this paper is to test alternative models of low-frequency variability and their parameters.

The plan of the paper is as follows. The next section introduces the three classes of models that we will consider: fractional models, local-to-unity autoregressions, and the local level model, parameterized as an unobserved components model with a large $I(0)$ component and a small unit root component. This section discusses the choice of weights for extracting the low-frequency components and the model-specific asymptotic distributions of the resulting weighted averages. Section 3 develops tests of the models based on these asymptotic distributions and studies their properties. Section 4 uses the methods of Section 3 to study the low-frequency properties of twenty macroeconomic and financial time series. Section 5 offers some additional comments on the feasibility of discriminating between the various low-frequency models.

2 Models and Low-Frequency Transformations

Let y_t, $t = 1, \cdots, T$ denote the observed time series, and consider the decomposition of y_t into unobserved deterministic and stochastic components

$$y_t = d_t + u_t. \tag{1}$$

This paper focuses on the low-frequency variability of the stochastic component u_t; the deterministic component is modelled as a constant $d_t = \mu$, or as a constant plus linear trend $d_t = \mu + \beta t$, with unknown parameters μ and β.

We consider five leading models used in finance and macroeconomics to model low-frequency variability. The first is a fractional (FR) or "long-memory" model; stationary versions of the model have a spectral density $S(\lambda) \propto |\lambda|^{-2d}$ as $\lambda \to 0$, where $-1/2 < d < 1/2$ is the fractional parameter. We follow Velasco (1999) and define the fractional model FR with $1/2 < d < 3/2$ for u_t when the first differences $u_t - u_{t-1}$ (with $u_0 = 0$) are a stationary fractional model with parameter $d - 1$. The second model is the autoregressive model with largest root close to unity; using standard notation we write the dominant autoregressive coefficient as $\rho_T = (1 - c/T)$, so that the process is characterized by the local-to-unity parameter c. For this model, normalized versions of u_t converge in distribution to an Ornstein-Uhlenbeck process with diffusion parameter $-c$, and for this reason we will refer to this as the OU model. We speak of the integrated OU model, I-OU, when $u_t - u_{t-1}$ (with $u_0 = 0$) follows the OU model. The forth model that we consider decomposes u_t into an

$I(0)$ and $I(1)$ component, $u_t = w_t + (g/T)\sum_{s=1}^{t}\eta_s$, where $(w_t, \eta_t)'$ are $I(0)$ with long-run covariance matrix $\sigma^2 I_2$, and g is a parameter that governs the relative importance of the $I(1)$ component. In this "Local Level" (LL) model (cf. Harvey (1989)) both components are important for the low-frequency variability of u_t. Again, we also define the integrated LL model, *I-LL,* as the model for u_t that arises when $u_t - u_{t-1}$ follows the LL model.

2.1 Asymptotic Representation of the Models

As shown below, the low-frequency variability implied by each of these models can be characterized by the stochastic properties of the partial sum process for u_t, so for our purposes it suffices to define each model in terms of the behavior of these partial sums of u_t. Table 1 summarizes the standard convergence properties of the partial sum process $T^{-\alpha}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \sigma G(\cdot)$ for each of the five models, where α is a model-specific constant and G is a model-specific mean-zero Gaussian process with covariance kernel $k(r,s)$ given in the final column of the table. A large number of primitive conditions have been used to justify these limits. Specifically, for the stationary fractional model, weak convergence to the fractional Wiener process W^d has been established under various primitive conditions for u_t by Taqqu (1975) and Chan and Terrin (1995)—see Marinucci and Robinson (1999) for additional references and discussion. The local-to-unity model and local level model rely on a functional central limit theorem applied to $(w_t, \eta_t)'$; various primitive conditions are given, for example, in McLeish (1974), Wooldridge and White (1988), Phillips and Solo (1992), and Davidson (2002); see Stock (1994) for general discussion.

The unit root and $I(0)$ models are nested in several of these models. The unit root model corresponds to the fractional model with $d = 1$, the OU model with $c = 0$, and the integrated local level model with $g = 0$. Similarly, the $I(0)$ model corresponds to the stationary fractional model with $d = 0$ and the local level model with $g = 0$.

The objective of this paper is to assess how well these specifications explain the low-frequency variability of the stochastic component u_t in (1). But u_t is not observed. We handle the unknown deterministic component d_t by restricting attention to statistics that are functions of the least-square residuals of a regression of y_t on a constant (denoted u_t^μ) or on a constant and time trend (denoted u_t^τ). Because $\{u_t^i\}_{t=1}^T$, $i = \mu, \tau$ are maximal invariants to the groups of transformations $\{y_t\}_{t=1}^T \to \{y_t + m\}_{t=1}^T$ and $\{y_t\}_{t=1}^T \to \{y_t + m + bt\}_{t=1}^T$, respectively, there is no loss of generality in basing inference on functions of $\{u_t^i\}_{t=1}^T$ for tests that are invariant to these transformations. Under the assumptions given above, a straightforward calculation shows that for $i = \mu, \tau$, $T^{-\alpha} \sum_{t=1}^{[\cdot T]} u_t^i \Rightarrow \sigma G^i(\cdot)$ where α is a model-specific constant and G^i is a model-specific mean-zero Gaussian process with covariance kernel $k^i(r,s)$ given by

$$
\begin{aligned}
k^\mu(r,s) &= k(r,s) - rk(1,s) - sk(r,1) + rsk(1,1) &(2)\\
k^\tau(r,s) &= k^\mu(r,s) - 6s(1-s)\textstyle\int k^\mu(r,\lambda)d\lambda &(3)\\
&\quad -6r(1-r)\textstyle\int k^\mu(\lambda,s)d\lambda + 36rs(1-s)(1-r)\int\int k^\mu(\nu,\lambda)d\nu d\lambda
\end{aligned}
$$

where $k(s,r)$ is the model's covariance kernel given in Table 1.

2.2 Asymptotic Properties of Weighted Averages

We extract the information about the low-frequency variability of u_t using a fixed number (q) of weighted averages of u_t^i, $i = \mu, \tau$, where the weights are known and deterministic low-frequency trigonometric series. We discuss and evaluate specific choices for the weight functions below, but first summarize the joint asymptotic distribution of these q weighted averages. Thus, let $\Psi : [0,1] \mapsto \mathbb{R}^q$ denote a set of q weight functions $\Psi = (\Psi_1, \cdots, \Psi_q)'$ with derivatives $\psi = (\psi_1, \cdots, \psi_q)'$, let $X_{Tj} = T^{-\alpha+1} \int_0^1 \Psi_j(s) u_{[sT]+1}^i ds = T^{-\alpha} \sum_{t=1}^T \tilde{\Psi}_{T,t,j} u_t^i$ with $\tilde{\Psi}_{T,t,j} = T \int_{(t-1)/T}^{t/T} \Psi_j(s)ds$ denote the $j'th$ weighted average, and $X_T = (X_{T1}, \cdots, X_{Tq})' = T^{-\alpha+1} \int_0^1 \Psi(s) u_{[sT]+1}^i ds$. If $T^{-\alpha} \sum_{t=1}^{[\cdot T]} u_t^i = G_T^i(\cdot) \Rightarrow \sigma G^i(\cdot)$, by integration by parts and the continuous mapping theorem,

$$
\begin{aligned}
X_T &= G_T^i(1)\Psi(1) - \textstyle\int_0^1 G_T^i(s)\psi(s)ds &(4)\\
&\Rightarrow X = -\sigma \textstyle\int_0^1 G^i(s)\psi(s)ds = -\sigma\int_0^1 \Psi(s)dG^i(s) \sim \mathcal{N}(0, \sigma^2\Sigma),
\end{aligned}
$$

since $G_T^i(1) = 0$. The covariance matrix Σ depends on the weight function Ψ and the covariance kernel for G^i, with j, lth element equal to $\int_0^1 \int_0^1 \psi_j(r)\psi_l(s)k^i(r,s)drds$ for $i = \mu, \tau$.

The convergence in distribution of X_T in (4) is an *implication* of the standard convergence $T^{-\alpha} \sum_{t=1}^{[\cdot T]} u_t^i \Rightarrow \sigma G^i(\cdot)$ for the 5 models discussed above. While, as a formal matter, (4) holds for any fixed value of q, it may provide a poor guide to the small sample behavior of X_T for a given sample size T if q is chosen very large. As an example, consider the case of a demeaned $I(0)$ model (so that G^{μ} is the demeaned Wiener process W^{μ} and $\alpha = 1/2$), and suppose $\Psi_j(s) = \sqrt{2}\cos(\pi js)$. As we show below, Σ in (4) then becomes $\Sigma = I_q$ leading to the asymptotic approximation of $\{X_{Tj}\}_{j=1}^{q}$ being i.i.d$\mathcal{N}(0,\sigma^2)$ for any fixed q. But $E[X_{Tj}^2]/(2\pi)$ is (almost) equal to the spectrum of u_t at frequency $j/2T$, so that this approximation implies a flat spectrum for frequencies below $q/2T$. Thus, for a given sample size (such as 60 years of quarterly data), it may make little sense to use approximations associated (4) for values of q that are large enough to incorporate business cycle (or higher) frequencies. Indeed, in this context, a reasonable *definition* for an $I(0)$ process (or any other of the 5 processes discussed above) in a macroeconomic context might thus be that (4) provides reasonable approximations for a choice of Ψ that captures below-business-cycle frequency variability.

If X_T captures the information in y_t about the low-frequency variability of u_t, then the question of model fit for a specific low-frequency model becomes the question whether X_T is approximately distributed $\mathcal{N}(0,\sigma^2\Sigma)$. For the models introduced above, Σ depends only on the model type and parameter value, so that $\Sigma = \Sigma_i(\theta)$ for $i \in \{FR, OU, I\text{-}OU, LL, I\text{-}LL\}$ and $\theta \in \{d, c, g\}$. The parameter σ^2 is an unknown constant governing the low-frequency scale of the process—for example, σ^2 is the long-run variance of η_t in the local-to-unity model. Because q is fixed (that is our asymptotics keep q fixed as $T \to \infty$) it is not possible to estimate σ^2 consistently using the q elements in X_T. This suggests restricting attention to scale invariant tests of X_T. Imposing scale invariance has the additional advantage that the value of α in $X_T = T^{-\alpha+1}\int_0^1 \Psi(s)u^i_{[sT]+1}ds$ does not need to be known.

Thus, consider the following maximal invariant to the group of transformation $X_T \to aX_T$, $a \neq 0$,

$$v_T = X_T/\sqrt{X_T'X_T}.$$

Under the conditions above, by the Continuous Mapping Theorem, $v_T \Rightarrow X/\sqrt{X'X}$. The density of $v = (v_1, \cdots, v_q)' = X/\sqrt{X'X}$ with respect to the uniform measure on the surface of a q dimensional unit sphere is given by (see, for instance, Kariya (1980) or King (1980))

$$f_v(\Sigma) = C|\Sigma|^{-1/2}\left(v'\Sigma^{-1}v\right)^{-q/2} \tag{5}$$

where the positive constant $C = \frac{1}{2}\Gamma(q/2)\pi^{-q/2}$, and $\Gamma(\cdot)$ is the Gamma function. For a given model for u_t, the asymptotic distribution of v_T depends only the $q \times q$ matrix $\Sigma_i(\theta)$, which is known for each model i and parameter θ. Our strategy therefore is to assess the model fit for a specific stochastic model i and parameter θ by testing whether v_T is approximately distributed (5) with $\Sigma = \Sigma_i(\theta)$.

2.3 Choice of Weights and the Resulting Covariance Matrices

Our choice of $\Psi = (\Psi_1, \cdots, \Psi_q)'$ is guided by two goals. The first goal is that Ψ should extract low-frequency variations of u_t and, to the extent possible, be uncontaminated by higher frequency variations. The second goal is that Ψ should produce a diagonal (or nearly diagonal) covariance matrix Σ, as this facilitates the interpretation of X_T because the models' implications for persistence in u_t become implications for specific forms of heteroskedasticity in X_T.

One way to investigate how well a candidate Ψ extracts low-frequency variability is to let u_t be exactly equal to a generic periodic series $u_t = \sin(\pi\vartheta t/T + \phi)$, where $\vartheta \geq 0$ and $\phi \in [0, \pi)$. The variability captured by X_T can then be measured by the R^2 of a regression of u_t on the demeaned/detrended weight functions. For T not too small, this R^2 is well approximated by the R^2 of a continuous time regression of $\sin(\pi\vartheta s + \phi)$ on $\Psi_1^i(s), \cdots, \Psi_q^i(s)$ on the unit interval, where Ψ_j^i, $i = \mu, \tau$, are the residuals of a continuous time regression of $\Psi_j(s)$ on 1 and $(1, s)$, respectively. Ideally, the R^2 should equal unity for $\vartheta \leq \vartheta_0$ and zero for $\vartheta > \vartheta_0$, for all phase shifts $\phi \in [0, \pi)$, where ϑ_0 corresponds to the pre-specified cut-off frequency. Standard sets of orthogonal trigonometric functions, such as the cosine expansion or the Fourier expansion with frequency smaller or equal to ϑ_0 are natural candidates for Ψ.

The left panels of Figure 1 plot R^2 as a function of ϑ for a cut-off frequency $\vartheta_0 = 14$ in the demeaned case, so that Ψ consists of the $q = 14$ elements[1] of the cosine expansion $\Psi_j(s) = \sqrt{2}\cos(\pi js)$ (denoted "eigenfunctions") and the Fourier expansion $\Psi_j(s) = \sqrt{2}\sin(\pi(j+1)s)$ for j odd and $\Psi_j(s) = \sqrt{2}\cos(\pi js)$ for j even. In the top panel, for each value of ϑ, R^2 is averaged over all values for the phase shift $\phi \in [0, \pi)$; in the middle panel, R^2 is maximized over ϕ and in the bottom panel, R^2 is minimized over ϕ. Both choices for Ψ_j come reasonably

[1] For postwar data in our empirical analysis, below business cycle variability is captured by $q = 13$ weighted averages in the demeaned case. We choose an even number here to ensure that in the demeaned case, the Fourier and cosine expansions have an equal number of elements.

close to the ideal of extracting all information about cycles of frequency $\vartheta \leq \vartheta_0$ ($R^2 = 1$) and no information about cycles of frequency $\vartheta > \vartheta_0$ ($R^2 = 0$).

In general, orthogonal functions Ψ_j^i only lead to a diagonal Σ in the $I(0)$ model, but not in persistent models—see, for instance, Akdi and Dickey (1998) for an analysis of the unit root model using the Fourier expansion. It is not possible to construct Ψ_j that lead to diagonal Σ for all models we consider. But consider a choice of Ψ_j as the eigenfunctions of the covariance kernel $k_W^\mu(r,s)$ and $k_W^\tau(r,s)$ of a demeaned and detrended Wiener process, respectively:

Theorem 1 *Let*

$$\begin{aligned} \varphi_j^\mu(s) &= \sqrt{2}\cos(\pi j s), \ \textit{for } j \geq 1 \\ \varphi_j^\tau(s) &= \begin{cases} \sqrt{2}\cos(\pi s(j+1)) & \textit{for odd } j \geq 1 \\ \sqrt{\frac{2\omega_{j/2}}{\omega_{j/2}-\sin(\omega_{j/2})}}(-1)^{(j+2)/2}\sin(\omega_{j/2}(s-1/2)) & \textit{for even } j \geq 2 \end{cases} \end{aligned}$$

$\varphi_0^\mu(s) = \varphi_{-1}^\tau(s) = 1$ *and* $\varphi_0^\tau(s) = \sqrt{3}(1-2s)$*, where* $\pi(2l+1) - \pi/6 < \omega_l < \pi(2l+1)$ *is the* l*th positive root of* $\cos(\omega/2) = 2\sin(\omega/2)/\omega$*. The set of orthonormal functions* $\{\varphi_j^\mu\}_{j=0}^\infty$ *and* $\{\varphi_j^\tau\}_{j=-1}^\infty$ *are the eigenfunctions of* $k_W^\mu(r,s)$ *and* $k_W^\tau(r,s)$ *with associated eigenvalues* $\{\lambda_j^\mu\}_{j=0}^\infty$ *and* $\{\lambda_j^\tau\}_{j=-1}^\infty$*, respectively, where* $\lambda_0^\mu = 0$ *and* $\lambda_j^\mu = (j\pi)^{-2}$ *for* $j \geq 1$ *and* $\lambda_{-1}^\tau = \lambda_0^\tau = 0$*,* $\lambda_j^\tau = (j\pi + \pi)^{-2}$ *for odd* $j \geq 1$ *and* $\lambda_j^\tau = (\omega_{j/2})^{-2}$ *for even* $j \geq 2$*.*

Theorem 1 identifies the cosine expansion $\sqrt{2}\cos(\pi j s)$, $j = 1, 2, \ldots$ as the eigenfunctions of $k_W^\mu(r,s)$ that correspond to nonzero eigenvalues, and also in the detrended case, the eigenfunctions $\varphi_j^\tau(s)$ are trigonometric functions. A natural choice for Ψ in the trend case are thus φ_j^τ, $j \geq 1$ with frequency smaller or equal ϑ_0. By construction, eigenfunctions result in a diagonal Σ for both the $I(1)$ and $I(0)$ models, and thus yield a diagonal Σ for all values of g in the local level model. For the fractional model and the OU model, the eigenfunctions produce a diagonal Σ only for $d = 0$ and $d = 1$, and for $c = 0$ and $c \to \infty$, respectively. Table 2 summarizes the size of the off-diagonal elements of Σ for various values of $\theta \in \{d, c, g\}$ in the FR, OU and LL models using the eigenfunctions. It presents the average absolute correlation when $\vartheta_0 = 14$, a typical value in the empirical analysis. The average absolute correlation is zero or close to zero for all considered parameter values.

What is more, the eigenfunctions φ_j^τ corresponding to nonzero eigenvalues are orthogonal to $(1, s)$, so that with $\Psi_j = \varphi_j^\tau$, the detrending to Ψ_j^τ leaves Ψ_j unaltered and thus orthogonal. This is not the case for the Fourier expansion. The choice of Ψ_j as the Fourier expansion of frequency smaller or equal ϑ_0 might thus inadvertently lead to more leakage of higher frequencies, as some linear combination of the *detrended* Fourier expansion approximates a higher frequency periodic series. This effect can be seen in the right panels of Figure 1, which contains R^2 plots for the eigenfunctions and the Fourier expansion in the detrended case with frequencies less than or equal to $\vartheta_0 = 14$ (so that $q = 13$ for the eigenfunctions and $q = 14$ for the Fourier expansion). We conclude that the eigenfunctions φ_j^i, $j = 1, 2, \ldots$ of Theorem 1 of frequency below the cut-off do a good job both at the extraction of low-frequency information with little leakage and at yielding approximately diagonal Σ for $i = \mu, \tau$, and the remainder of the paper is based on this choice.

With this choice, the covariance matrix Σ is close to diagonal, so the models can usefully be compared by considering the diagonal elements of Σ only. Figure 2 plots the square roots of these diagonal elements for the various models considered in Table 2. Evidently, more persistent models produce larger variances for low-frequency components, a generalization of the familiar 'periodogram' intuition that for stationary u_t, the variance of $\sqrt{2/T}\sum_{t=1}^{T}\cos(\pi jt/T)u_t$ is approximately equal to 2π times the spectral density at frequency $j/2T$. For example, for the unit root model ($d = 1$ in the fractional model or $c = 0$ in the OU model), the standard deviation of X_1 is 14 times larger than the standard deviation of X_{14}. In contrast, when $d = 0.25$ in the fractional model the relative standard deviation of X_1 falls to 1.8, and when $c = 5$ in the OU model, the relative standard deviation of X_1 is 6.3. In the $I(0)$ model ($d = 0$ in the fractional model or $g = 0$ in the local level model), $\Sigma = I_q$, and all of the standard deviations are unity.

2.4 Continuity of the Fractional and Local-to-Unity Models

It is useful to briefly discuss the continuity of $\Sigma_i(\theta)$ for two of the models. In the local-to-unity model, there is a discontinuity at $c = 0$ in our treatment of the initial condition and this leads to different covariance kernels in Table 1; similarly, in the fractional model there is a discontinuity at $d = 1/2$ as we move from the stationary to the integrated version of the model. As it turns out, these discontinuities do not lead to discontinuities of the density of v in (5) as a function of c and d.

This is easily seen in the local-to-unity model. Location invariance implies that it suffices to consider the asymptotic distribution of $T^{-1/2}(u_{[\cdot T]} - u_1)$. As noted by Elliott (1999), in the stable model $T^{-1/2}(u_{[\cdot T]} - u_1) \Rightarrow J^c(\cdot) - J^c(0) = Z(e^{-sc} - 1)/\sqrt{2c} + \int_0^s e^{-c(s-\lambda)} dW(\lambda)$, and $\lim_{c \downarrow 0}(e^{-sc} - 1)/\sqrt{2c} = 0$, so that the asymptotic distribution of $T^{-1/2}(u_{[\cdot T]} - u_1)$ is continuous at $c = 0$.

The calculation for the fractional model is somewhat more involved. Note that the density (5) of v remains unchanged under reparametrizations $\Sigma \to a\Sigma$ for any $a > 0$. Because $\Sigma_{FR}(d)$ is a linear function of $k^i(r,s)$, it therefore suffices to show that

$$\lim_{\epsilon \downarrow 0} \frac{k^i_{FR(\frac{1}{2}-\epsilon)}(r,s)}{k^i_{I\text{-}FR(\frac{1}{2}+\epsilon)}(r,s)} = a \tag{6}$$

for some constant $a > 0$ that does not depend on (r,s), where $k^i_{FR(d)}$ and $k^i_{I\text{-}FR(d)}$ are the covariance kernels for the stationary and integrated fractional models with parameter d for $i = \mu, \tau$. As shown in the appendix, (6) holds with $a = 2$, so that the density of v is continuous at $d = 1/2$.[2]

3 Test Statistics

This section discusses several test statistics for the models. As discussed above, when (4) holds, the transformed data satisfies $v_T \Rightarrow v = X/\sqrt{X'X}$ with $X \sim \mathcal{N}(0, \Sigma)$. The low-frequency characteristics of the models are summarized by the covariance matrix $\Sigma = \Sigma_i(\theta)$, which is known for a given model $i \in \{FR, OU, I\text{-}OU, LL, I\text{-}LL\}$ and model parameter θ. A test of adequacy of a given model and parameter value can therefore be conducted by testing

[2] This result suggests a definition of a demeaned or detrended fractional process with $d = 1/2$ as any process whose partial sums converge to a Gaussian process with covariance kernel that is given by an appropriately scaled limit of k^{μ}_{FR} or k^{τ}_{FR} as $d \uparrow 1/2$; see equations (12) and (13) in the appendix. The possibility of a continuous extension across all values of d renders Velasco's (1999) definition of fractional processes with $d \in (1/2, 3/2)$ as the partial sums of a stationary fractional process with parameter $d - 1$ considerably more attractive, as it does not lead to a discontinuity at the boundary $d = 1/2$, at least for demeaned or detrended data with appropriately chosen scale.

$H_0 : \Sigma = \Sigma_0$ against $H_1 : \Sigma \neq \Sigma_0$. This section derives optimal tests for this problem based on v (or, equivalently, optimal scale invariant tests based on X). Because there exists no uniformly powerful test, one must take a stand for what kind of alternatives tests should be powerful. We consider four optimal tests that direct power to different alternatives. The first two tests are low-frequency versions of point-optimal unit root and "stationarity" tests: these tests focus on two specific null hypotheses (the $I(1)$ and the $I(0)$ models) and maximize power against the local-to-unity and local level models, respectively. The final two tests are relevant for any null model; the first maximizes weighted average power against alternatives that correspond to misspecification of the persistence in u_t, and the second maximizes weighted average power against alternatives that correspond to misspecification of the second moment of u_t. For all four tests, we follow King (1988) and choose the distance from the null so that a 5% level test has approximately 50% power at the alternative for which it is optimal.

The tests we derive are optimal scale invariant tests based on X, the limiting random variable in $X_T \Rightarrow X$. As shown by Müller (2007a), these tests, when applied to X_T (i.e. v_T), are optimal in the sense that they maximize (weighted average) power among all scale invariant tests whose asymptotic rejection probability is smaller or equal to the nominal level for *all* data generating processes satisfying $X_T \Rightarrow X \sim \mathcal{N}(0, \Sigma_0)$. In other words, if the convergence $X_T \Rightarrow X$ of (4) completely summarizes the implications for data y_t generated by a given low-frequency model, then the test statistics derived in this section applied to v_T are asymptotically most powerful (in a weighted average sense) among all scale invariant asymptotically valid tests.

3.1 Low-Frequency I(1) and I(0) tests

We test the $I(1)$ and $I(0)$ null hypotheses using low-frequency point-optimal tests. Specifically, in the context of the local-to-unity model we test the unit root model $c = c_0 = 0$ against the alternative model with $c = c_1$ using the likelihood ratio statistic

$$\text{LFUR} = v'\Sigma_{OU}(c_0)^{-1}v/v'\Sigma_{OU}(c_1)^{-1}v$$

where the value of c_1 is chosen so that the 5%-level test has power of approximately 50% when $c = c_1$ for the model with $q = 13$ (a typical value in our empirical analysis). This yields $c_1 = 14$ for demeaned series and $c_1 = 28$ for detrended series. We label the statistic LFUR as a reminder that it is a low-frequency unit root test statistic.

We similarly test the $I(0)$ null hypothesis against the point alternative of a local level model with parameter $g = g_1 > 0$ (which is the same nesting of the $I(0)$ model as employed in Nyblom (1989) and Kwiatkowski, Phillips, Schmidt, and Shin (1992)). A calculation shows that the likelihood ratio statistic rejects for large values of

$$\text{LFST} = \left(\sum_{j=1}^{q} v_j^2\right) / \left(\sum_{j=1}^{q} \frac{v_j^2}{1 + g_1^2 \lambda_j}\right)$$

where λ_j are the eigenvalues defined in Theorem 1. The 50% power requirement, imposed for $q = 13$, yields approximately $g_1 = 10$ in the mean case and $g_1 = 20$ in the trend case.

3.2 Testing for Misspecified Persistence in u_t

As discussed in Section 2, low-frequency persistence in u_t leads to heteroskedasticity in X, so that that misspecification of the persistence for u_t translates into misspecification of the heteroskedasticity function for X. This motivates a specification test that focuses on the diagonal elements of Σ. Thus, let Λ denote a diagonal matrix, and consider an alternative of the form $\Sigma = \Lambda \Sigma_0 \Lambda$. The relative magnitudes of the diagonal elements of Λ distort the relative magnitude of the diagonal elements of Σ_0, and produce values of Σ associated with processes that are more or less persistent than the null model. For example, decreasing diagonal elements of Λ represent u_t with more persistence (more very low frequency variability) than under the null model. More complicated patterns for the diagonal elements of Λ allow more subtle deviations from the null model in the persistence features of u_t.

To detect a variety of departures from the null, we consider several different values of Λ and construct a test with best weighted average power over these alternatives. Letting F denote the weight function for Λ, the best test is simply the Neyman-Pearson test associated with a null in which v has density $f_v(\Sigma_0)$ and an alternative in which the density of v is the F-weighted mixture of $f_v(\Lambda \Sigma_0 \Lambda)$. The details of the test involve the choice of values of Λ and their associated weights.

A simple and flexible way to specify the values of Λ and corresponding weights F is to represent Λ as $\Lambda = \text{diag}(\exp(\delta_1), \cdots, \exp(\delta_q))$, where $\delta = (\delta_1, \cdots, \delta_q)'$ is a mean zero Gaussian vector with covariance matrix $\gamma^2\Omega$. Specifically, the empirical analysis has δ_j follow a random walk: $\delta_j = \delta_{j-1} + \varepsilon_j$ with $\delta_0 = 0$ and $\varepsilon_j \sim iid\mathcal{N}(0, \gamma^2)$. For this choice, the weighted average power maximizing test seeks to detect misspecification in the exact form of persistence in u_t against a wide range of alternatives, while maintaining that the implied heteroskedasticity in X is relatively smooth. The weighted average power maximizing test is the best test of the simple hypotheses

$$H_0 : v \text{ has density } f_v(\Sigma_0) \quad \text{vs.} \quad H_1 : v \text{ has density } E_\delta f_v(\Lambda\Sigma_0\Lambda) \tag{7}$$

where E_δ denotes integration over the measure of δ and f_v is defined in (5). By the Neyman-Pearson Lemma and the form of f_v, an optimal test of (7) rejects for large values of

$$S = \frac{E_\delta[|\Lambda\Sigma_0\Lambda|^{-1/2}(v'(\Lambda\Sigma_0\Lambda)^{-1}v)^{-q/2}]}{(v'\Sigma_0^{-1}v)^{-q/2}}.$$

Because the power of the S test does not depend on Σ_0 when Σ_0 is diagonal (which is approximately true for all of the models considered), the same value of γ^2 can be used to satisfy the 50% power requirement for all models for a given q. Numerical analysis shows $\gamma = 5/q$ to be a good choice for values of q ranging from 5 to 30.

3.3 Testing for Low-Frequency Heteroskedasticity in u_t

Limiting results for partial sums like those shown in Table 1 are robust to time varying variances of the driving disturbances, as long as the time variation is a stationary short memory process; this implies that the values of Σ are similarly robust to such forms of heteroskedasticity. However, instability in the second moment of financial and macroeconomic data is often quite persistent (e.g., Bollerslev, Engle, and Nelson (1994) and Andersen, Bollerslev, Christoffersen, and Diebold (2007), Balke and Gordon (1989), Kim and Nelson (1999) and McConnell and Perez-Quiros (2000)), so it is interesting to ask whether second moments of u_t exhibit enough low-frequency variability to invalidate limits like those shown in Table 1. To investigate this, we nest each of the models considered thus far in a more general model that allows for such low-frequency heteroskedasticity, derive the resulting value of Σ for the more general model, and construct an optimal test against such alternatives.

For each of the low-frequency models, we consider a version of the model with low-frequency heteroskedastic driving disturbances in their natural moving average representations. For example, for the $I(0)$ model consider models for $\{u_t\}$ that satisfy $T^{-1/2}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \sigma\int_0^{\cdot} h(\lambda)dW_1(\lambda)$, where $h : [0,1] \mapsto \mathbb{R}$ is a continuous function. When $h(s) = 1$ for all s this yields the $I(0)$ model in Table 1, but non-constant h yield different limiting processes and different values of Σ. Indeed, a calculation shows that the $lj'th$ element of Σ is $\Sigma_{lj} = \int_0^1 \Psi_l(s)\Psi_j(s)h(s)^2 ds$. Because the Ψ functions are orthogonal, $\Sigma_{lj} = 0$ for $l \neq j$ when h is constant, but non-constant h lead to non-zero values of Σ_{lj}. Said differently, low-frequency heteroskedasticity in u_t leads to serial correlation in X. The form of this serial correlation depends on h. For example, in the mean case when $h(s) = \sqrt{1 + 2a\cos(\pi s)}$ with $|a| < 1/2$, Σ turns out to equal the covariance matrix of a MA(1) process with first-order autocorrelation equal to a.

The same device can be used to generalize the other models. Thus, consider $\{u_t\}$ that satisfy $T^{-\alpha}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \sigma\tilde{G}(\cdot)$, where for the different models:

$$\begin{array}{rcl}
FR & : & \left\{\begin{array}{l}
\tilde{G}(s) = A(d)\int_{-\infty}^{0}((s-\lambda)^d - (-\lambda)^d)dW(\lambda) \\
\qquad\qquad\qquad +A(d)\int_0^s (s-\lambda)^d h(\lambda)dW(\lambda),\ d \in (-1/2, 1/2) \\
\tilde{G}(s) = \frac{A(d-1)}{d}\int_{-\infty}^{0}((s-\lambda)^d - (-\lambda)^{d-1}(sd-\lambda))dW(\lambda) \\
\qquad\qquad\qquad +\frac{A(d-1)}{d}\int_0^s (s-\lambda)^d h(\lambda)dW(\lambda),\ d \in (1/2, 3/2)
\end{array}\right. \\
OU & : & \left\{\begin{array}{l}
\tilde{G}(s) = \frac{1}{c}\int_{-\infty}^{0}(e^{c\lambda} - e^{-c(s-\lambda)})dW(\lambda) + \frac{1}{c}\int_0^s (1 - e^{-c(s-\lambda)})h(\lambda)dW(\lambda),\ c > 0 \\
\tilde{G}(s) = \int_0^s (s-\lambda)h(\lambda)dW(\lambda),\ c = 0
\end{array}\right. \\
LL & : & \tilde{G}(s) = \int_0^s h(\lambda)dW_1(\lambda) + g\int_0^s (s-\lambda)h(\lambda)dW_2(\lambda),\ g \geq 0 \\
I\text{-}OU & : & \begin{array}{l}
\tilde{G}(s) = c^{-2}\int_{-\infty}^{0}(e^{-c(s-\lambda)} - (1-cs)e^{c\lambda})dW(\lambda) \\
\qquad\qquad\qquad +c^{-2}\int_0^s (e^{-(s-\lambda)} - c(s-\lambda) - 1)h(\lambda)dW(\lambda),\ c > 0
\end{array} \\
I\text{-}LL & : & \tilde{G}(s) = \int_0^s (s-\lambda)h(\lambda)dW_1(\lambda) + \frac{1}{2}g\int_0^s (s-\lambda)^2 h(\lambda)dW_2(\lambda),\ g \geq 0
\end{array} \tag{8}$$

In these representations the function h only affects the stochastic component of $\tilde{G}(s)$ that stems from the in-sample innovations, but leaves unaffected terms associated with initial conditions, such as $\frac{1}{c}\int_{-\infty}^{0}(e^{-c(s-\lambda)} - e^{c\lambda})dW(\lambda)$ in the stable local-to-unity model. The idea is that $h(t/T)$ describes the square root of the time varying long-run variance of the in-sample driving disturbances at date $t \geq 1$, while maintaining the assumption that stable models were stationary prior to the beginning of the sample. This restriction means that $\tilde{G}(s)$ is a sum of two pieces, and the one that captures the pre-sample innovations remains unaffected by h. Especially in the fractional model, such a decomposition is computationally convenient, as noted by Davidson and Hashimadze (2006). As in the $I(0)$ example, for any of the models and any continuous function h, it is possible to compute the covariance kernel for $\tilde{G}$, and the resulting covariance matrix of X.

Let $\Sigma_i(\theta_0, h)$ denote the value of Σ associated with model i with parameter θ_0 and heteroskedasticity function h. The homoskedastic versions of the models from Table 1 then yield $\Sigma = \Sigma_i(\theta_0, 1)$ while their heteroskedastic counterparts yield $\Sigma = \Sigma_i(\theta_0, h)$. The goal therefore is to look for departures from the null hypothesis $\Sigma = \Sigma_i(\theta_0, 1)$ in the direction of alternatives of the form $\Sigma = \Sigma_i(\theta_0, h)$. Because there is no uniformly most powerful test over all functions h, we consider a test with best weighted average power for a wide range of h functions. The details of the test involve the choice of values of h and their associated weights.

Similar to the choice of values of Λ for the S test, we consider a flexible model for the values of h that arise as realizations from a Wiener process. In particular, we consider functions generated as $h = e^{\eta W^*}$, where W^* is a standard Wiener process on the unit interval independent of G, and η is a parameter. The test with best weighted average power over this set of h functions is the best test associated with the hypotheses

$$H_0 : v \text{ has density } f_v(\Sigma(\theta_0, 1)) \quad \text{vs.} \quad H_1 : v \text{ has density } E_{W^*} f_v(\Sigma(\theta_0, e^{\eta W^*})) \tag{9}$$

where E_{W^*} denotes integration over the distribution of W^*. The form of f_v and the Neyman-Pearson Lemma imply that the optimal test of (9) rejects for large values of

$$H = \frac{E_{W^*}[|\Sigma(\theta_0, e^{\eta W^*})|^{-1/2}(v'\Sigma(\theta_0, e^{\eta W^*})^{-1}v)^{-q/2}]}{(v'\Sigma(\theta_0, 1)^{-1}v)^{-q/2}}.$$

A choice of $\eta = 6q^{-1/2}$ satisfies the 50% power requirement for a wide range of values of q for both the $I(0)$ and $I(1)$ models.

3.4 Some Properties of the Tests

This section takes up the issues of the asymptotic power of the various tests for various alternatives and the accuracy of the asymptotic approximations in finite samples. The numerical results will be presented for 5%-level tests, $q = 13$ and demeaned data.

3.4.1 Asymptotic Power

The asymptotic rejection frequency of the LFUR and LFST tests is shown in Figure 3 for a range of values of d in the fractional model (panel (a)), c in the OU model (panel (b)), and g in local level model (panel (c)). For example, panel (b) shows that the LFST test has power of approximately 90% for the unit root alternative ($c = 0$), but power less of less than 25% for values of c greater than 20 in the OU model. Applying this asymptotic approximation to an autoregression using $T = 200$ observations, the LFST test will reject the $I(0)$ null with high probability when the largest AR root is unity, but is unlikely to reject the $I(0)$ null when the largest root is less than $1 - 20/200 = 0.9$. Similarly, from panel (c), the LFUR test has power of over 90% for the $I(0)$ model, but power of less than 25% for values of g greater then 20. This asymptotic approximation suggests that in a MA model for $(1 - L)y_t$ and with $T = 200$ observations, the LFUR test will reject the $I(1)$ null with high probability when the MA polynomial has a unit root (that is, when the level of y_t is $I(0)$) , but is unlikely to reject when the largest MA root is less than 0.9.

Figure 3 also allows power comparisons between the optimal low-frequency tests and tests that use all frequencies. For example, from panel (b), the $q = 13$ low-frequency unit root test has approximately 50% power when $c = 14$. This is the best power that can be achieved when y_t exhibits low-frequency OU behavior with $c = 0$ under the null. If instead, y_t followed an exact Gaussian AR(1) model with unit AR coefficient under the null and local-to-unity coefficient under the alternative, then it would be appropriate to use all frequencies to test the null, and the best all-frequency test has approximately 75% power when $c = 14$ (cf. Elliott (1999)). This 25% difference in power is associated with the relatively weaker restriction on the null model of the LFUR test, with an assumption that the I(1) model only provides an accurate description of low-frequency behavior, while allowing for unrestricted behavior of the series at higher frequencies.

Figure 4 compares the power of the S test to the power of the LFUR and LFST tests, with alternatives of the form $\Sigma_1 = \Lambda\Sigma_0\Lambda$ where $\Lambda = \text{diag}(\exp(\delta_1), \cdots, \exp(\delta_q))$. Because Λ is diagonal, the power of the S tests does not depend on Σ_0 when Σ_0 is diagonal, and because Σ_0 is exactly or approximately diagonal for all of the models considered in Table 2, the power results for the S apply to each of these models. In contrast, the LFST and LFUR tests utilize particular values of Σ_0, so the results for LFST apply to the $I(0)$ null and the result for LFUR apply to the $I(1)$ null. In panel (a) δ_i follows the linear trend $\delta_i = \kappa(i-1)/(q-1)$, where $\kappa = 0$ yields $\Sigma_1 = \Sigma_0$, $\kappa < 0$ produces models with more persistence than the null model, and $\kappa > 0$ produces models with less persistence. In panel (b), $\{\delta_i\}$ has a triangular shape: $\delta_i = \kappa(i-1)/6$ for $i \leq 7$, and $\delta_i = \delta_{14-i}$ for $8 \leq i \leq 13$. As in panel (a), $\kappa = 0$ yields $\Sigma_1 = \Sigma_0$, but now non-zero values of κ correspond to non-monotonic deviations in the persistence of u_t across frequencies. Because the LFST test looks for alternatives that are more persistent than the null hypothesis, it acts as a one-sided test for $\kappa < 0$ in panel (a) and it has power less than size when $\kappa > 0$. Similarly, LFUR acts as one-sided test for alternatives that are less persistent than the null, and is biased when $\kappa < 0$. In contrast, the S test looks for departures from the null in several directions (associated with realizations from draws a demeaned random walk), and panel (a) indicates that it is approximately unbiased with a symmetric power function that is roughly comparable to the one-sided LFST and LFUR tests under this alternative. Panel (b), which considers the triangular alternative, shows a power function for S that is similar to the trend alternative, while the power functions for LFST and LFUR indicate bias and (because of the non-monotonicity of the alternative) these tests have one-sided power that is substantially less than the one-sided power for the trend alternative shown in panel (a).

Figure 5 presents the power of the H test. Because low-frequency heteroskedasticity in u_t leads to serial correlation in X, we compare the power of the H tests to two tests for serial correlation in X: let $\rho_i = \left(\sum_{j=1}^{q-i} X_j X_{j+i}\right) / \left(\sum_{j=1}^{q} X_j^2\right)$; the first test statistic is $|\rho_1|$ (and thus checks for first order serial correlation), while the second is $\sum_{i=1}^{q-1} |\rho_i|/i$ (and checks for serial correlation at all lags). Figure 5 shows results for the $I(0)$ null model where $\ln(h(s))$ follows a linear trend model in panel (a) ($\ln(h(s)) = \kappa s$) and a triangular model in panel (b) ($\ln(h(s)) = \kappa s$ for $s \leq 1/2$ and $\ln(h(s)) = \kappa(1-s)$ for $s > 1/2$). In panel (a), the power of H is slightly smaller than the power of the $|\rho_1|$ test for values of κ near zero, slightly larger for more distant alternatives, and the $|\rho_1|$ test appears to dominate the $\sum_{i=1}^{q-1}|\rho_i|/i$ test. All of the tests are approximately unbiased with symmetric power functions. In panel (b), where the alternative involves a non-monotonic heteroskedasticity function $h(s)$, the H test remains approximately unbiased with a power function that is symmetric, but the two other tests are biased and show better power performance for $\kappa < 0$.

3.4.2 Finite Sample Performance

There are three distinct issues related to the finite sample performance of the tests. First, the data used in the tests (X_T) are weighted averages of the original data (y_t), and by virtue of the central limit theorem the probability distribution of X_T is approximated by the normal distribution. Second, as we implement the tests in the empirical section below, the covariance matrix of X_T is approximated by the covariance matrix of X, that is by the expression below equation (4). Finally, our analysis is predicated on the behavior of the process over a set of low-frequencies, but as the R^2 functions shown in Figure 1 indicate, there is some contamination in X_T caused by leakage from higher frequencies. The first of these issues—the quality of the normal approximation to the distribution of a sample average—is well-studied, and we say nothing more about it except to offer the reminder that because X_T is a weighted average of the underlying data, it is exactly normally distributed when the underlying data y_t are normal. As for the second issue—the approximation associated with using the asymptotic form of the covariance matrix for X_T—we have investigated the quality of the approximation for empirically relevant values of T, and found it to be very good. For example, using $T = 200$, $q = 13$, and i.i.d. Gaussian data, the size of the asymptotic 5%-level LFUR test is 0.05, and the power for the AR(1) model $(1 - 0.95L)y_t = \varepsilon_t$ is 0.36, which can be compared to the asymptotic power for $c = 10$, which is 0.35.

To investigate the third issue—leakage of high frequency variability into X_T—consider two experiments. In the first experiment stationary Gaussian data are generated from a stochastic process with spectrum $s(\omega) = 1$ for $|\omega| \leq 2\pi/R$ and $s(\omega) = \kappa$ for $|\omega| > 2\pi/R$, where R is a cut-off period used to demarcate low-frequency variability. When $\kappa = 1$ the spectrum is flat, but when $\kappa \neq 1$, the spectrum is step function with discontinuity at

$|\omega| = 2\pi/R$. With $R = 32$ and $T = 208$ this corresponds to 52 years of quarterly data with a cut-off frequency corresponding to a period of 8 years and implies that $q = 13$ (and the choice with T/R a natural number maximizes potential leakage). Since the spectrum is constant for low frequencies independent of the value of κ, one would want the small sample rejection probability of LFST, S and H tests under the $I(0)$ null hypothesis to be equal to the nominal level. A second experiment uses partial sums of the data from the first experiment to compute the small sample rejection probability of LFUR, S, and H tests under the $I(1)$ null. In both experiments, size distortions of 5% level tests based on asymptotic critical values are small for $0 \leq \kappa \leq 3$: The largest size is 7.1% for the S test in the $I(1)$ model, and the smallest is 3.6% for the LFST test in the $I(0)$ model, both with $\kappa = 3$. These experiments suggest that leakage is not a serious concern.

4 Empirical Results

In this section we use the low-frequency tests to address four empirical questions. The first is the Nelson-Plosser question: after accounting for a deterministic linear trend, is real GDP consistent with the $I(1)$ model? The second is a question about the cointegration of long term and short term interest rates: is the term spread consistent with the $I(0)$ model? We answer both of these questions using post-war quarterly U.S. data and focus the analysis on periods greater than 32 periods (that is, frequencies lower than the business cycle). The third question involves the behavior of real exchange rates where a large literature has commented on the connection between the persistence of real exchange rates and deviations from purchasing power parity. Here we ask whether a long annual series on real exchange rates is consistent with the $I(0)$ model over *any* set of low frequencies, and this allows us to construct a confidence set for the range of low frequencies consistent with the $I(0)$ model. Finally, we use the S and H tests to construct confidence sets for the parameters of the five low-frequency models for below-business-cycle variability in twenty U.S. macroeconomic and financial time series.

4.1 Testing the I(0) and I(1) Null for Real GDP and the Term Spread

Table 3 shows selected empirical results for quarterly values (1952:1-2005:3) of the logarithm of (detrended) real GDP and the (demeaned) term spread—the difference between interest rates for 10 year and 1 year U.S. Treasury bonds. Panel (a) shows results computed using standard methods: p-values for the DFGLS unit root test of Elliott, Rothenberg and Stock (1996), the stationarity test of Nyblom (1989) (using a HAC covariance matrix as suggested in Kwiatkowski, Phillips, Schmidt, and Shin (1992)), and the estimated values of d and standard errors from Geweke and Porter-Hudak (1983) or "GPH regressions" as described in the Robinson (2003). Panel (b) shows p-values for the LFST, LFUR, S, and H tests under the $I(0)$ and $I(1)$ nulls.

Looking first at the results for real GDP, the traditional statistics shown in panel (a) suggest that the data are consistent with the $I(1)$ model, but not the $I(0)$ model: the p-value for the DFGLS test 0.16, while the Nyblom/KPSS test has a p-value less than 1%; the GPH regressions produce point estimates of d close to the unit root null, and the implied confidence intervals for d include the $I(1)$ model but exclude the $I(0)$ model. The low-frequency results shown in panel (b) reach the same conclusion: the p-value for the LFUR test is 37% while the p-value for the LFST test is less than 1%. Moreover, the H statistic indicates that the well known post-1983 decline in volatility of real GDP (the "Great Moderation") is not severe enough to reject the homoskedastic low-frequency $I(1)$ model. Thus for real GDP, the low-frequency inference and inference based on traditional methods largely coincide.

This is not the case for the term spread. Panel (a) indicates that the DFGLS statistic rejects the $I(1)$ model, the Nyblom/KPSS tests rejects the $I(0)$ model, and the results from the GPH regressions depend critically on whether $[T^{0.5}]$ or $[T^{0.65}]$ observations are used in the GPH regression. In contrast, panel (b) shows that the low-frequency variability in the series is consistent with the $I(0)$ model but not with the $I(1)$ model. Thus, traditional inference methods paint a murky picture, while the low-frequency variability of the series is consistent with the hypothesis that long rates and short rates are cointegrated.

4.2 A Low-Frequency Confidence Interval for I(0) Variability in Real Exchange Rates

A large empirical literature has examined the unit root or near unit root behavior of real exchange rates. The data used here—annual observations on the real dollar/pound real exchange rate from 1791-2004—come in large part from one important empirical study in this literature, Lothian and Taylor (1996). A natural question to ask in the context of the framework developed here is whether real exchange rates are consistent with $I(0)$ behavior over *any* low-frequency band. That is, is there any value of q such that the low-frequency transformed data X_T is consistent with the $I(0)$ model, and more generally what is largest value of q (or the highest frequency) consistent with the $I(0)$ model? Figure 6 plots the p-value of the LFST test applied to the logarithm of the (demeaned) real exchange rate for values of $q \leq 30$, corresponding to periods longer than approximately 14 years. The figure shows that $I(0)$ model is rejected at the 5% level for values of $q > 7$ (periods shorter than 61 years) and at the 1% level for values of $q > 10$ (periods shorter than 43 years). Equivalently, a 95% confidence interval for periods for which the real exchange rate behaves like an $I(0)$ process includes periods greater 61 years, while a 99% confidence interval includes periods grates than 43 years. In either case, the results suggest very long departures from purchasing power parity.

4.3 Confidence Intervals for Model Parameters

Confidence sets for model parameters can be formed by inverting the S and H tests. Table 4 presents the resulting confidence sets for twenty macroeconomic and financial time series that include post-war quarterly versions of important macroeconomic aggregates (real GDP, aggregate inflation, nominal and real interest rates, productivity, and employment) and longer annual versions of related series (real GNP from 1869-2004, nominal and real bond yields from 1900-2004, and so forth). We also study several cointegrating relations by analyzing differences between series (such as the long-short interest rate spread discussed above) or logarithms of ratios (such as consumption-income or dividend-price ratios). A detailed description of the data is given in the Appendix. As usual, several of the data series are transformed by taking logarithms, and as discussed above, the deterministic component of each series is modeled as a constant or a linear trend. Table A.1 summarizes these transformations for each series. The post-war quarterly series span the period 1952:1-2005:3, so that $T = 215$, and $q = [2T/32] = 13$ for the demeaned series and $q = 12$ for the detrended series. Each annual time series is available for a different sample period (real GNP is available from 1869-2004, while bond rates are available from 1900-2004, real exchange rates from 1791-2004, for example), so the value of q is series-specific. One series (returns on the SP500) contains daily observations from 1928-2005, and for this series $q = 17$.

Looking at the table, it is clear that the relatively short sample (less than 60 years of data for many of the series), means that confidence sets often contain a wide ranges of values for d, c, and g. That said, looking at the individual series, several results are noteworthy:

The unit root model for inflation is not rejected using the post-war quarterly data, while the $I(0)$ model is rejected. Results are shown for inflation based on the GDP deflator, but similar conclusions follow from the PCE deflator and CPI. Stock and Watson (2007) document instability in the "size" of the unit root component (corresponding to the value of g in the local level model) over the post-war period, but apparently this instability is not so severe that it leads to rejections based on the tests considered here. Different results are obtained from the long-annual (1869-2004) series, which shows less persistence than the postwar quarterly series.

Labor productivity is very persistent. The $I(0)$ model is rejected but the $I(1)$ model is not. The S test rejects values of $c > 5$ in the OU model and $d < 0.84$ in the fractional model. The behavior of employee hours per capita has received considerable attention in the recent VAR literature (see Gali (1999), Christiano, Eichenbaum, and Vigfusson (2003), Pesavento and Rossi (2005), and Francis and Ramey (2005)). The results shown here are consistent with unit root but not $I(0)$ low-frequency behavior. Francis and Ramey (2006) discuss demographic trends that are potentially responsible for the high degree of persistence in this series.

Postwar nominal interest rates are consistent with a unit root but not an $I(0)$ process, and the S statistic similarly rejects the $I(0)$ model of the long annual nominal bond rates. In contrast, the $I(0)$ model is not rejected using the S test for real interest rates.

Several of the data series, such as the logarithm of the ratio of consumption to income, represent error correction terms from putative cointegrating relationships. Under the hypothesis of cointegration, these series should be $I(0)$. Table 4 shows that real unit labor costs (the logarithm of the ratio of labor productivity to real wages, $y-n-w$ in familiar notation) exhibit limited persistence: the $I(1)$ model is rejected but the $I(0)$ model is not rejected. The "balanced growth" cointegrating relation between consumption and income (e.g., King, Plosser, Stock, and Watson (1991)) fares less well, where the $I(1)$ model is not rejected, but the $I(0)$ model is rejected. The apparent source of this rejection is the large increase in the consumption-income ratio over the 1985-2004 period, a subject that has attracted much recent attention (for example, see Lettau and Ludvigson (2004) for an explanation based on increases in asset values.) The investment-income relationship also appears to be at odds with the null of cointegration (although, in results not reported in the table, this rejection depends in part on the particular series used for investment and its deflator.) Finally, the stability of the logarithm of the earnings-stock price ratio or dividend-price ratio, and the implication of this stability for the predictability of stock prices, has been an ongoing subject of controversy (see Campbell and Yogo (2006) for a recent discussion). Using Campbell and Yogo's (2006) annual earnings/price data for the SP500 from 1880-2002, both the $I(0)$ and $I(1)$ models are rejected (and similar results are found for their dividend/price data over the same sample period). Confidence intervals constructed using the S statistic suggest less persistence than a unit root (for example the confidence interval for the fractional model includes $0.28 \leq d \leq 0.90$). The shorter (1928-2004) CRSP dividend-yield (also from Campbell and Yogo (2006)), displays more low-frequency persistence, and is consistent with the $I(1)$ model but not the $I(0)$ model.

Ding, Granger, and Engle (1993) analyzed the absolute value of daily returns from the SP500 and showed that the autocorrelations decayed in a way that was remarkably consistent with a fractional process. Low frequency characteristics of the data are consistent with this finding. Both the unit root and $I(0)$ models are rejected by the S statistic, but models with somewhat less persistence than the unit root, such as the fraction model with $0.08 < d < 0.88$, are not rejected.

A striking finding from Table 4 is the number of models that are rejected by the H test.

For example, the low-frequency heteroskedasticity in the long annual GDP series (in large part associated with the post-WWII decline in volatility) leads to a rejection for *all* of the models. Several of the other series show similar rejections or produce confidence intervals for model parameters with little overlap with the confidence intervals associated with the S tests. These results suggest that for many of the series, low-frequency heteroskedasticity is so severe that it invalidates the limiting results shown in Table 1.

Two main findings stand out from these empirical analysis. First, despite focus on the narrow below business-cycle frequency band, very few of the series are compatible with the $I(0)$ model. This hold true even for putative cointegration error correction terms involving consumption, income, and investment, and stock prices and earnings. Most macroeconomic series and relationships thus exhibit pronounced non-trivial dynamics below business cycle frequencies. In contrast, the unit root model is often consistent with the observed low-frequency variability.

Second, maybe the most important empirical conclusion is that for many series there seems to be too much low-frequency variability in the second moment to provide good fits for any of the models. From an economic perspective, this underlines the importance of understanding the sources and implications of such low-frequency volatility changes. From a statistical perspective, this finding motivates further research into methods that allow for substantial time variation in second moments.

5 Additional Remarks

The analysis in this paper has focused on tests for whether a low-frequency model with a specific parameter value is a plausible data generating mechanism for the transformed data v_T. Alternatively, one might ask whether a model as such, with unspecified parameter value, is rejected in favor of another model. A large number of inference procedures have been developed for specific low-frequency models, such as the local-to-unity model and the fractional model. Yet, typically there is considerable uncertainty about the appropriate low-frequency model for a given series. A high-power discrimination procedure would therefore have obvious practical appeal.

Consider then the problem of discriminating between the three continuous bridges between the $I(0)$ and the $I(1)$ model: the fractional model with $0 \leq d \leq 1$, the local-to-unity model with $c \geq 0$ and the local level model with $g \geq 0$. These models are obviously similar in the sense that they all nest (or arbitrarily well approximate) the $I(0)$ and $I(1)$ model. More interestingly, a recent literature has pointed out that (non-degenerate) regime switching models and fractional models are similar along many dimensions—see, for example, Parke (1999), Diebold and Inoue (2001), and Davidson and Sibbertsen (2005). Since the local level model can be viewed as a short memory model with time varying mean, this question is closely related to the similarity of the fractional model with $0 < d < 1$ and the local level model with $g > 0$.

This suggests that it will be challenging to discriminate between low-frequency models using information contained in v_T. A convenient way to quantify the difficulty is to compute the total variation distance between the models. Recall that the total variation distance between two probability measures is defined as the largest absolute difference the two probability measures assign to the same event, maximized over all events. Let Σ_0 and Σ_1 be the covariance matrices of X induced by two models and specific parameter values. Using a standard equality (see, for instance, Pollard (2002), page 60), the total variation distance between the two probability measures described by the densities $f_v(\Sigma_0)$ and $f_v(\Sigma_1)$ is given by

$$\mathrm{TVD}(\Sigma_0, \Sigma_1) = \tfrac{1}{2} \int |f_v(\Sigma_0) - f_v(\Sigma_1)| d\eta(v)$$

where η is the uniform measure on the surface of a q dimensional unit sphere. There is no obvious way to analytically solve this integral, but it can be evaluated using Monte Carlo integration. To see how, write

$$\begin{aligned} \mathrm{TVD}(\Sigma_0, \Sigma_1) &= \int \mathbf{1}[f_v(\Sigma_1) < f_v(\Sigma_0)](f_v(\Sigma_0) - f_v(\Sigma_1)) d\eta(v) \\ &= \int \mathbf{1}[\mathrm{LR}_v < 1](1 - \mathrm{LR}_v) f_v(\Sigma_0) d\eta(v) \qquad (10) \end{aligned}$$

where $\mathrm{LR}_v = f_v(\Sigma_1)/f_v(\Sigma_0)$. Thus, $\mathrm{TVD}(\Sigma_0, \Sigma_1)$ can be approximated by drawing v's under $f_v(\Sigma_0)$ and averaging the resulting values of $\mathbf{1}[\mathrm{LR}_v < 1](1 - \mathrm{LR}_v)$.[3]

[3] It is numerically advantageous to rely on (10) rather than on the more straightforward expression

Let $\Sigma_i(\theta)$ denote the covariance matrix of X for model $i \in \{FR, OU, LL\}$ with parameter value θ, and consider the quantity

$$D_{i,j}(\theta) = \min_{\gamma \in \Gamma} \mathrm{TVD}(\Sigma_i(\theta), \Sigma_j(\gamma))$$

where $\Gamma = [0,1]$ for $j = FR$ and $\Gamma = [0,\infty)$ for $j \in \{OU, LL\}$. If $D_{i,j}(\theta)$ is small, then there is a parameter value $\gamma_0 \in \Gamma$ for which the distribution of v with $\Sigma = \Sigma_j(\gamma_0)$ is close to the distribution of v with $\Sigma = \Sigma_i(\theta)$, so it will be difficult to discriminate model i from model j if indeed $\Sigma = \Sigma_i(\theta)$. More formally, consider any model discrimination procedure between models i and j based on v, which correctly chooses model i when $\Sigma = \Sigma_i(\theta)$ with probability p. By definition of the total variation distance, the probability of the event "procedure selects model i" under $\Sigma = \Sigma_j(\gamma)$ is at least $p - \mathrm{TVD}(\Sigma_i(\theta), \Sigma_j(\gamma))$. If $D_{i,j}(\theta)$ is small, then either the probability of mistakenly selecting model i is large for some $\Sigma = \Sigma_j(\gamma_0)$, $\gamma_0 \in \Gamma$, or the probability p of correctly selecting model i is small. In the language of hypothesis tests, for any test of the the null hypothesis that $\Sigma = \Sigma_j(\gamma)$, $\gamma \in \Gamma$ against the alternative that $\Sigma = \Sigma_i(\theta)$, $\theta \in \Theta$, the sum of the probabilities of Type I and Type II error are bounded below by $1 - \max_{\theta \in \Theta} D_{i,j}(\theta)$.

The value of $D_{i,j}(\theta)$ is an (increasing) function of q. Figure 7 plots $D_{i,j}(\theta)$ for each of the model pairs for $q = 13$, which corresponds to 52 years of data with interest focused on frequencies lower than 8-year cycles. Panel (a) plots $D_{FR,OU}(d)$ and $D_{FR,LL}(d)$ and panels (b) and (c) contain similar plots for the OU and LL models. Evidently $D_{i,j}(\theta)$ is small throughout. For example, for all values of d, the largest distance of the fractional model to the local-to-unity and local level model is less than 25%, and the largest distance between the OU and LL models is less than 45%. For comparison, the total variation distance between the $I(0)$ and $I(1)$ model for $q = 13$ is approximately 90%. Total variation distance using detrended data is somewhat smaller than the values shown in Figure 7. Evidently then, it is impossible to discriminate between these standard models with any reasonable level of confidence using sample sizes typical in macroeconomic applications, at least based on the below business cycle variability in the series summarized by v_T. Indeed, to obtain, say,

$\mathrm{TVD}(\Sigma_0, \Sigma_1) = \frac{1}{2}\int |1 - \mathrm{LR}_v| f_v(\Sigma_0) d\eta(v)$ for the numerical integration, since $\mathbf{1}[\mathrm{LR}_v < 1](1 - \mathrm{LR}_v)$ is bounded and thus possesses all moments, which is not necessarily true for $|1 - \mathrm{LR}_v|$.

$\max_{0 \leq d \leq 1} D_{FR,OU}(d) \approx 0.9$, one would need a sample size of 480 years (corresponding to $q = 120$).

These results imply that it is essentially impossible to discriminate between these models based on low-frequency information using sample sizes typically encountered in empirical work. When using any one of these one parameter low-frequency models for empirical work, one thus must either rely on extraneous information to argue for the correct model choice, or one must take these models seriously over a much wider frequency band. Neither of these two options is particularly attractive for many applications, which raises the question whether econometric techniques can be developed that remain valid for a wide range of low-frequency models.

A Appendix

A.1 Proof of Theorem 1

Applying (2) and (3) to $k_W(r,s) = E[W(r)W(s)] = \min(r,s)$ yields

$$\begin{aligned} k_W^\mu(r,s) &= \min(r,s) + \tfrac{1}{3} - (r+s) + \tfrac{1}{2}(r^2+s^2) \\ k_W^\tau(r,s) &= \min(s,r) + \sum_{l=1}^{4} \varsigma_{5-l}(r)\varsigma_l(s) \end{aligned} \tag{11}$$

where $\varsigma_1(r) = \frac{1}{15} - \frac{11}{10}r + 2r^2 - r^3$, $\varsigma_2(r) = \frac{3}{5}r - 3r^2 + 2r^3$, $\varsigma_3(r) = r$ and $\varsigma_4(r) = 1$. Noting that for any real $\lambda \neq 0$, $s > 0$ and ϕ

$$\begin{aligned} \int_0^s \sin(\lambda u + \phi) u du &= (\sin(\lambda s + \phi) - \lambda s \cos(\lambda s + \phi) - \sin(\phi))/\lambda^2 \\ \int_0^s \sin(\lambda u + \phi) u^2 du &= (2s\lambda \sin(\lambda s + \phi) + (2 - \lambda^2 s^2)\cos(\lambda s + \phi) - 2\cos(\phi))/\lambda^3 \\ \int_0^s \sin(\lambda u + \phi) u^3 du &= (3(\lambda^2 s^2 - 2)\sin(\lambda s + \phi) + \lambda s(6 - \lambda^2 s^2)\cos(\lambda s + \phi) + 6\sin(\phi))/\lambda^4 \end{aligned}$$

it is straightforward, but very tedious, to confirm that $\int_0^1 k_W^i(r,s)\varphi_j^i(s)ds = \lambda_j^i \varphi_j^i(r)$ for $j = 0, 1, \ldots$ when $i = \mu$ and for $j = -1, 0, 1, 2, \ldots$ when $i = \tau$.

Note that $\{\varphi_j^\mu\}_{j=0}^\infty$ is necessarily the complete set of eigenfunctions, since the cosine expansion is a basis of $L^2[0,1]$. For the detrended case, it is not hard to see that the two functions φ_{-1}^τ and φ_0^τ are the only possible eigenfunctions of $k_W^\tau(r,s)$ that correspond to a zero eigenvalue. Furthermore, Nabeya and Tanaka (1988) show that eigenfunctions of kernels of the form (11) corresponding to nonzero eigenvalues, i.e. functions f satisfying $\int_0^1 k_W^\tau(r,s)f(s)ds = \lambda f(r)$ with $\lambda \neq 0$, are the solutions of the second order differential equation $f''(s) + \lambda f(s) = \sum_{l=1}^2 a_l \varsigma_l''(s)$ under some appropriate boundary conditions. Since ς_1'' and ς_2'' are linear, we conclude that f is of the form $f(s) = c_1 \cos(\sqrt{\lambda}s) + c_2 \sin(\sqrt{\lambda}s) + c_3 + c_4 s$. It thus suffices to show that $\int_0^1 f(s)\varphi_j^\tau(s)ds = 0$ for $j \geq -1$ implies $c_l = 0$ for $l = 1, \cdots, 4$. As $\varphi_{-1}^\tau(s)$ and $\varphi_0^\tau(s)$ span $\{1, s\}$, and φ_j^τ, $j \geq 1$ are orthogonal to φ_{-1}^τ and φ_0^τ, this is equivalent to showing $\int_0^1 f(s)\varphi_j^\tau(s)ds = 0$ for $j \geq 1$ implies $c = 0$ in the parametrization $f(s) = c\sin(\omega(s - \frac{1}{2}) + \phi)$, $\omega > 0$ and $\phi \in (-\pi, \pi)$. A calculations yields that $\int_0^1 f(s)\varphi_1^\tau(s)ds = 0$ and $\int_0^1 f(s)\varphi_{2[\omega/2\pi]-1}^\tau(s)ds = 0$ imply $\phi = 0$ or $c = 0$, and $c\int_0^1 \sin(\omega(s - \frac{1}{2}))\varphi_{2[\omega/2\pi]}^\tau(s)ds = 0$ implies $c = 0$.

A.2 Continuity of fractional process at $d = 1/2$:

By the definition of $k^{\mu}_{FR(d)}(r,s)$ and $k^{\mu}_{I\text{-}FR(d)}(r,s)$, we find for $s \leq r$

$$
\begin{aligned}
k^{\mu}_{FR(d)}(r,s) = \tfrac{1}{2}[s^{1+2d} + r^{1+2d} - (r-s)^{1+2d} + 2rs \\
- s(1-(1-r)^{1+2d} + r^{1+2d}) - r(1-(1-s)^{1+2d} + s^{1+2d})]
\end{aligned}
$$

and

$$
\begin{aligned}
k^{\mu}_{I\text{-}FR(d)}(r,s) = \frac{1}{4d(1+2d)}[-r^{1+2d}(1-s) - s(s^{2d} + (r-s)^{2d} + (1-r)^{2d} - 1) \\
+ r(s^{1+2d} + 1 - (1-s)^{2d} + (r-s)^{2d}) + sr((1-s)^{2d} + (1-r)^{2d} - 2))]
\end{aligned}
$$

so that

$$
\lim_{d \uparrow 1/2} k^{\mu}_{FR(d)}(r,s) = \lim_{d \downarrow 1/2} k^{\mu}_{I\text{-}FR(d)}(r,s) = 0.
$$

Now for $0 < s < r$, using that for any real $a > 0$, $\lim_{x \downarrow 0}(a^x - 1)/x = \ln a$, we find

$$
\begin{aligned}
\lim_{d \uparrow 1/2} \frac{k^{\mu}_{FR(d)}(r,s)}{1/2-d} = -(1-r)^2 s \ln(1-r) - r^2(1-s)\ln r - r(1-s)^2 \ln(1-s) \\
+(r-s)^2 \ln(r-s) + (r-1)s^2 \ln s
\end{aligned}
\tag{12}
$$

and

$$
\lim_{d \uparrow 1/2} \frac{k^{\mu}_{FR(d)}(r,r)}{1/2-d} = 2(1-r)r(-(1-r)\ln(1-r) - r\ln r). \tag{13}
$$

Performing the same computation for $\lim_{d \downarrow 1/2} k^{\mu}_{I\text{-}FR(d)}(r,s)/(d-1/2)$ yields the desired result in the demeaned case. The detrended case follows from these results and (3).

A.3 Data Appendix

Table A1 lists the series used in Section 4, the sample period, data frequency transformation, and data source and notes.

References

Akdi, Y., and D. A. Dickey (1998): "Periodograms of Unit Root Time Series: Distributions and Tests," *Communications in Statistics: Theory and Methods*, 27, 69–87.

Andersen, T., T. Bollerslev, P. Christoffersen, and F. X. Diebold (2007): *Volatility: Practical Methods for Financial Applications*. Princeton University Press, forthcoming, Princeton.

Balke, N. S., and R. J. Gordon (1989): "The Estimation of Prewar Gross National Product: Methodology and New Evidence," *Journal of Political Economy*, 94, 38–92.

Beveridge, S., and C. R. Nelson (1981): "A New Approach to Decomposition of Economics Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 7, 151–174.

Bierens, H. J. (1997): "Nonparametric Cointegration Analysis," *Journal of Econometrics*, 77, 379–404.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994): "ARCH Models," in *Handbook of Econometrics Vol. IV*, ed. by R. F. Engle, and D. McFadden. Elsevier Science, Amsterdam.

Campbell, J. Y., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60.

Chan, N. H., and N. Terrin (1995): "Inference for Unstable Long-Memory Processes with Applications to Fractional Unit Root Autoregressions," *Annals of Statistics*, 23, 1662–1683.

Christiano, L., M. Eichenbaum, and R. Vigfusson (2003): "What Happens After a Technology Shock," *NBER Working Paper 9819.*

Davidson, J. (2002): "Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes," *Journal of Econometrics*, 106, 243–269.

DAVIDSON, J., AND N. HASHIMADZE (2006): "Type I and Type II Fractional Brownian Motions: A Reconsideration," *Working paper, University of Exeter.*

DAVIDSON, J., AND P. SIBBERTSEN (2005): "Generating Schemes for Long Memory Processes: Regimes, Aggregation and Linearity," *Journal of Econometrics*, 128, 253–282.

DIEBOLD, F. X., AND A. INOUE (2001): "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131–159.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–116.

ELLIOTT, G. (1999): "Efficient Tests for a Unit Root When the Initial Observation is Drawn From its Unconditional Distribution," *International Economic Review*, 40, 767–783.

FAMA, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25, 383–417.

FRANCIS, N., AND V. A. RAMEY (2005): "Is the Technology-Driven Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited," *Journal of Monetary Economics*, 52, 1379–1399.

——— (2006): "Measures of Per Capita Hours and their Implications for the Technology-Hours Debate," *Working paper, U.C. San Diego.*

GALI, J. (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, 89, 249–271.

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

HARVEY, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

KARIYA, T. (1980): "Locally Robust Test for Serial Correlation in Least Squares Regression," *Annals of Statistics*, 8, 1065–1070.

KIM, C.-J., AND C. R. NELSON (1999): "Has the Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics*, 81, 608–616.

KING, M. L. (1980): "Robust Tests for Spherical Symmetry and their Application to Least Squares Regression," *The Annals of Statistics*, 8, 1265–1271.

——— (1988): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218.

KING, R., C. I. PLOSSER, J. H. STOCK, AND M. W. WATSON (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–840.

KWIATKOWSKI, D., P. C. B. PHILLIPS, P. SCHMIDT, AND Y. SHIN (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159–178.

LETTAU, M., AND S. C. LUDVIGSON (2004): "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption," *American Economic Review*, 94, 276–299.

LOTHIAN, J. R., AND M. P. TAYLOR (1996): "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries," *Journal of Political Economy*, 104, 488–509.

MARINUCCI, D., AND P. M. ROBINSON (1999): "Alternative Forms of Fractional Brownian Motion," *Journal of Statistical Planning and Inference*, 80, 111–122.

MCCONNELL, M. M., AND G. PEREZ-QUIROS (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980's," *American Economic Review*, 90, 1464–1476.

MCLEISH, D. L. (1974): "Dependent Central Limit Theorems and Invariance Principles," *The Annals of Probability*, 2, 620–628.

MEESE, R. A., AND K. ROGOFF (1983): "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?," *Journal of International Economics*, 14, 3–24.

MÜLLER, U. K. (2007a): "An Alternative Sense of Asymptotic Efficiency," *Working paper, Princeton University.*

——— (2007b): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141, 1331–1352.

NABEYA, S., AND K. TANAKA (1988): "Asymptotic Theory of a Test for Constancy of Regression Coefficients Against the Random Walk Alternative," *Annals of Statistics*, 16, 218–235.

NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series — Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

NYBLOM, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

PARKE, W. R. (1999): "What is Fractional Integration?," *Review of Economics and Statistics*, 81, 632–638.

PESAVENTO, E., AND B. ROSSI (2005): "Do Technology Shocks Drive Hours Up or Down? A Little Evidence from an Agnostic Procedure," *Macroeconomic Dynamics*, 9, 478–488.

PHILLIPS, P. C. B. (1998): "New Tools for Understanding Spurious Regression," *Econometrica*, 66, 1299–1325.

——— (2006): "Optimal Estimation of Cointegrated Systems with Irrelevant Instruments," *Cowles Foundation Discussion Paper 1547.*

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

POLLARD, D. (2002): *A User's Guide to Measure Theoretic Probability.* Cambridge University Press, Cambridge, UK.

Robinson, P. M. (2003): "Long-Memory Time Series," in *Time Series with Long Memory*, ed. by P. M. Robinson, pp. 4–32. Oxford University Press, Oxford.

Said, S. E., and D. A. Dickey (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 2599–607.

Stock, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4, pp. 2740–2841. North Holland, New York.

Stock, J. H., and M. W. Watson (2007): "Why Has Inflation Become Harder to Forecast?," *Journal of Money, Credit, and Banking*, 39, 3–34.

Taqqu, M. S. (1975): "Convergence of Integrated Processes of Arbitrary Hermite Rank," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50, 53–83.

Velasco, C. (1999): "Non-Stationary Log-Periodogram Regression," *Journal of Econometrics*, 91, 325–371.

Wooldridge, J. M., and H. White (1988): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210–230.

Table A1
Data Description and Sources

Series	Sample Period	F	Tr	Source and Notes
Real GDP	1952:1-2005:3	Q	ln τ	DRI: GDP157
Real GNP (Long Annual)	1869-2004	A	ln τ	1869-1928: Balke and Gordon (1989) 1929-2004: BEA (Series are linked in 1929)
Inflation	1952:1-2005:3	Q	lev μ	DRI: 400×ln(GDP272(t)/GDP272(t−1))
Inflation (Long Annual)	1870-2004	A	lev μ	GNP Deflator (PGNP): 1869-1928: Balke and Gordon (1989) 1929-2004: BEA (Series are linked in 1929) Inflation Series is 100×ln(PGNP(t)/PGNP(t−1))
Productivity	1952:1-2005:2	Q	ln τ	DRI: LBOUT (Output per hour, business sector)
Hours	1952:1-2005:2	Q	ln τ	DRI: LBMN(t)/P16(t) (Employee hours/population)
10YrTBond	1952:1-2005:3	Q	lev μ	DRI: FYGT10
1YrTBond	1952:1-2005:3	Q	lev μ	DRI: FYGT1
3mthTbill	1952:1-2005:2	Q	lev μ	DRI:FYGM3
Bond Rate	1900-2004	A	lev μ	NBER: M13108 (1900-1946) DRI: FYAAAI (1947-2004)
Real Tbill Rate	1952:1-2005:2	Q	lev μ	DRI: FYGM3(t)-400×ln(GDP273(t+1)/GDP273(t))
Real Bond Rate	1900-2004	A	lev μ	R(t) – 100×ln(PGNP(t)/PGNP(t−1)) R(t) = Bond Rate (described above) PGNP = GNP deflator (described above)
Dollar/Pound Real Ex. Rate	1791-2004	A	ln μ	1791-1990: Lothian and Taylor (1996) 1991-2004: FRB (Nominal Exchange Rate) BLS (US PPI Finished Goods) IFS (UK PPI Manufactured Goods)
Unit Labor Cost	1952:1-2005:2	Q	ln μ	DRI: LBLCP(t)/LBGDP(t)
TBond Spread	1952:1-2005:3	Q	lev μ	DRI: FYGT10-FYGT1
real C-GDP	1952:1-2005:3	Q	lnrμ	DRI: GDP 158/GDP157
real I-GDP	1952:1-005:3	Q	lnrμ	DRI: GDP 177/ GDP 157
Earnings/Price (SP500)	1880-2002	A	lnrμ	Campbell and Yogo (2006)
Div/Price (CRSP)	1926-2004	A	lnrμ	Campbell and Yogo (2006)
Abs.Returns (SP500)	1/3/1928-1/22/2005	D	lnrμ	SP: SP500(t) is the closing price at date t. Absolute returns are |ln[SP500(t)/SP500(t−1)]|

Notes: The column labeled *F* shows the data frequency (*A*: annual, *Q*: quarterly, and *D*: daily). The column labeled *Tr* (transformation) show the transformation: demeaned levels (lev μ), detrended levels (lev τ), demeaned logarithms (ln μ), detrended logarithms (ln τ), and lnr denotes the logarithm of the indicated ratio. In the column labeled *Source and Notes*, DRI denotes the DRI Economics Database (formerly Citibase) and NBER denotes the NBER historical data base.

Table 1
Asymptotic Properties of Partial Sums of Popular Time Series Models

Process	**Parameter**	**Partial sum convergence**	**Covariance kernel $k(r,s)$, $s \le r$**
1a. FR	$-\frac{1}{2} < d < \frac{1}{2}$	$T^{-1/2-d}\sigma^{-1}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow W^d(\cdot)$	$\frac{1}{2}(r^{2d+1} + s^{2d+1} - (r-s)^{2d+1})$
1b. FR	$\frac{1}{2} < d < \frac{3}{2}$	$T^{-1/2-d}\sigma^{-1}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \int_0^{\bullet} W^{d-1}(l)dl$	$\dfrac{(r-s)^{2d+1} + (1+2d)(rs^{2d} + r^{2d}s) - r^{2d+1} - s^{2d+1}}{4d(1+2d)}$
2a. OU	$c > 0$	$T^{-3/2}\sigma^{-1}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \int_0^{\bullet} J^c(l)dl$	$\dfrac{2cs - 1 + e^{-cs} + e^{-cr} - e^{-c(r-s)}}{2c^3}$
2b. OU	$c = 0$	$T^{-3/2}\sigma^{-1}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \int_0^{\bullet} W(l)dl$	$\frac{1}{6}(3rs^2 - s^3)$
3. I-OU	$c > 0$	$T^{-5/2}\sigma^{-1}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \int_0^{\bullet}\int_0^{\lambda} J^c(l)dld\lambda$	$\dfrac{3 - sc(3 + c^2s^2) + 3rc(1 - cs + c^2s^2) - 3e^{-cs}(1+cr) - 3e^{-cr}(1 + cs - e^{cs})}{6c^5}$
4. LL	$g \ge 0$	$T^{-1/2}\sigma^{-1}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow W_1(\cdot) + g\int_0^{\bullet} W_2(l)dl$	$s + \frac{1}{6}g^2(3rs^2 - s^3)$
5. I-LL	$g \ge 0$	$T^{-3/2}\sigma^{-1}\sum_{t=1}^{[\cdot T]} u_t \Rightarrow \int_0^{\bullet} W_1(l)dl + g\int_0^{\bullet}\int_0^{\lambda} W_2(l)dld\lambda$	$\frac{1}{6}(3rs^2 - s^3) + \frac{1}{120}g^2(10r^2s^3 - 5rs^4 + s^5)$

Notes: W, W_1 and W_2 are independent standard Wiener processes, W^d is "type I" fractional Brownian Motion defined as $W^d(s) = A(d)\int_{-\infty}^{0}\left[(s-l)^d - (-l)^d\right]dW(l) + A(d)\int_0^s (s-l)^d\, dW(l)$ where $A(d) = \left(\frac{1}{2d+1} + \int_0^{\infty}\left[(1+l)^d - l^d\right]^2 dl\right)^{-1/2}$ and J^c is the stationary Ornstein-Uhlenbeck process $J^c(s) = Ze^{-sc}/\sqrt{2c} + \int_0^s e^{-c(s-l)}dW(l)$ with $Z \sim N(0,1)$ independent of W.

Table 2
Average Absolute Correlations for $\Sigma(\theta)$

Fractional Model	***d* = −0.25**	***d* = 0.00**	***d* = 0.25**	***d* = 0.75**	***d* = 1.00**	***d* = 1.25**
Demeaned	0.03	0.00	0.01	0.01	0.00	0.03
Detrended	0.03	0.00	0.01	0.01	0.00	0.02
OU Model	***c* = 30**	***c* = 20**	***c* = 15**	***c* = 10**	***c* = 5**	***c* = 0**
Demeaned	0.02	0.02	0.02	0.02	0.02	0.00
Detrended	0.02	0.02	0.02	0.02	0.01	0.00
Local Level Model	***g* = 0**	***g* = 2**	***g* = 5**	***g* = 10**	***g* = 20**	***g* = 30**
Demeaned	0.00	0.00	0.00	0.00	0.00	0.00
Detrended	0.00	0.00	0.00	0.00	0.00	0.00

Notes: Entries in the table are the average values of the absolute values of the correlations associated with $\Sigma(\theta)$ with q = 14 for the demeaned model and q = 13 for the detrended model.

Table 3: Results Real GDP and Term Spread

A. DFGLS, Nyblom/KPSS and GPH Results

Series	DFGLS *p*-value	Nyblom/ KPSS *p*-value	GPH Regressions: $\hat{d}$ (SE)			
			Levels		Differences	
			$[T^{0.5}]$	$[T^{0.65}]$	$[T^{0.5}]$	$T^{0.65}$
Real GDP	0.16	<0.01	1.00 (0.17)	0.98 (0.11)	-0.19 (0.17)	-0.09 (0.11)
TBond Spread	<0.01	0.01	0.18 (0.17)	0.61 (0.11)	-0.80 (0.17)	-0.41 (0.11)

B. *P*-values for *I*(0) and *I*(1) Models

Series	I(0)			I(1)		
	LFST	S	H	LFUR	S	H
Real GDP	<0.01	0.09	0.12	0.37	0.84	0.59
TBond Spread	0.25	0.46	0.14	<0.01	<0.01	0.05

Notes: The entries in the column labeled *DFGLS* are *p*-values for the DFGLS test of Elliott, Rothenberg and Stock (1996). The entries in the column labeled *Nyblom* are *p*-values for the Nyblom (1989) *I*(0) test (using a HAC covariance matrix as suggested in Kwiatkowski, Phillips, Schmidt and Shin (1992)). Results are computed using a Newey-West HAC estimator with $[0.75\times T^{1/3}]$ lags. The results in the columns labeled *GPH Regressions* are the estimated values of *d* and standard errors computed from regressions using the lowest $[T^{0.5}]$ or $[T^{0.65}]$ periodogram ordinates. The GPH regressions and standard errors were computed as described in the Robinson (2003): specifically, the *GPH* regressions are of the form $ln(p_i) = \beta_0 + \beta_1 ln(\omega_i)$ + error, where p_i is the i'th periodogram ordinate and ω_i is the corresponding frequency, the estimated value of $\hat{d} = -\hat{\beta}_1 / 2$, where $\hat{\beta}_1$ is the OLS estimator, and the standard error of $\hat{d}$ is $SE(\hat{d}) = \pi/\sqrt{24m}$, where m is the number of periodogram ordinates used in the regression.

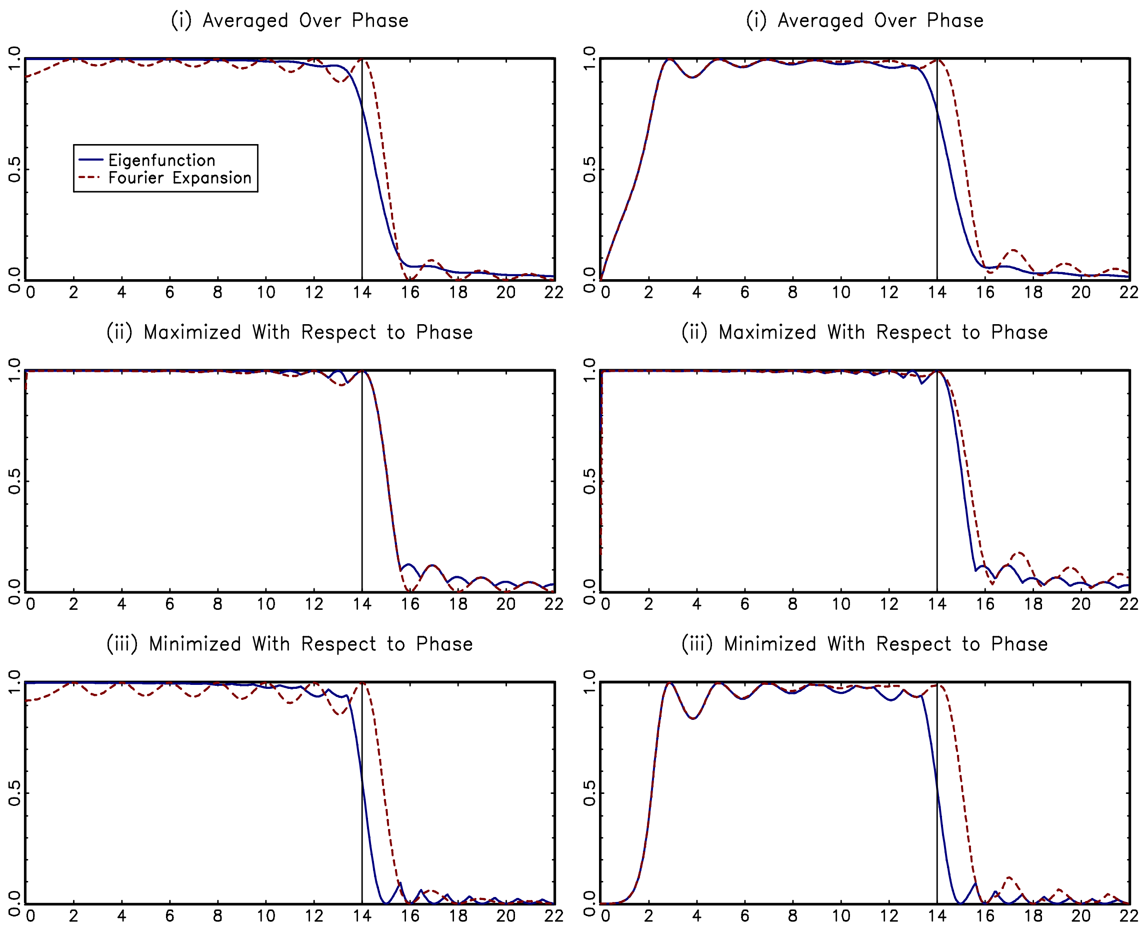
Table 4: 95% Confidence Sets for Model Parameters using *S* and *H* tests

Series	*Fractional Model (d)*		OU Model (*c*)		LL Model (*g*)		Integrated OU Model (*c*)		Integrated LL Model (*g*)	
	S	***H***	***S***	***H***	***S***	***H***	***S***	***H***	***S***	***H***
Post-War Quarterly Time Series										
Real GDP	(-0.06,1.50)	(-0.14,1.50)	(0.0,30.0)	(0.0,30.0)	(0.0,30.0)	(0.0,30.0)	(6.0,30.0)	(0.0,30.0)	(0.0,30.0)	(0.0,30.0)
TBond Spread	(-0.44,0.54)	(-0.50,0.96)	(15.0,30.0)	(0.5,30.0)	(0.0,16.5)	(0.0,30.0)				
Inflation	(0.26,1.48)	(-0.50,-0.30) (0.26,1.50)	(0.0,27.5)	(0.0,28.5)	(7.5,30.0)	(3.5,30.0)	(12.5,30.0)	(0.0,30.0)	(0.0,27.0)	(0.0,30.0)
Productivity	(0.84,1.50)	(-0.30,1.50)	(0.0,5.0)	(0.0,30.0)		(0.0,30.0)	(0.0,30.0)	(0.0,30.0)	(0.0,30.0)	(0.0,30.0)
Hours	(0.06,1.50)	(-0.50,1.50)	(0.0,30.0)	(0.0,30.0)	(2.5,30.0)	(0.0,30.0)	(10.0,30.0)	(0.0,30.0)	(0.0,28.5)	(0.0,30.0)
10YrTBond	(0.44,1.48)	(-0.50,0.34) (0.48,1.08)	(0.0,15.0)	(0.0,0.5) (12.5,30.0)	(10.5,30.0)	(0.0,30.0)	(6.5,30.0)	(30.0,30.0)	(0.0,30.0)	(0.0,0.5)
1YrTBond	(0.22,1.30)	(-0.50,1.42)	(0.0,30.0)	(0.0,30.0)	(6.0,30.0)	(0.0,30.0)	(29.5,30.0)	(0.0,0.0) (6.0,30.0)	(0.0,10.5)	(0.0,30.0)
3mthTbill	(0.18,1.28)	(-0.50,1.24)	(0.0,30.0)	(0.0,30.0)	(5.0,30.0)	(0.0,30.0)	(29.0,30.0)	(17.0,30.0)	(0.0,10.0)	(0.0,2.5)
Real Tbill Rate	(-0.24,1.48)	(-0.50,1.38)	(0.0,30.0)	(0.0,4.0) (12.0,30.0)	(0.0,30.0)	(0.0,30.0)	(10.5,30.0)	(19.0,30.0)	(0.0,24.0)	(0.0,2.5)
Unit Labor Cost	(-0.12,0.76)	(-0.50,1.50)	(6.5,30.0)	(0.0,30.0)	(0.0,28.0)	(0.0,30.0)		(0.0,30.0)		(0.0,30.0)
real C-GDP	(0.52,1.30)	(-0.50,0.26) (0.40,1.50)	(0.0,9.0)	(0.0,10.5) (17.5,30.0)	(8.0,30.0)	(0.0,30.0)	(9.5,30.0)	(0.0,30.0)	(0.0,8.0)	(0.0,30.0)
real I-GDP	(0.38,1.34)	(-0.42,0.34)	(0.0,13.0)		(5.5,30.0)	(0.0,17.0)	(17.5,30.0)		(0.0,13.0)	
Long Annual Time Series										
Real GNP	(0.24,1.18)		(0.0,30.0)		(19.5,30.0)				(0.0,16.5)	
Inflation	(0.06,0.60)	(-0.50,-0.06) (1.50,1.50)	(29.0,30.0)		(2.5,30.0)			(0.0,3.0)		
Real Ex. Rate	(0.52,1.04)	(-0.50,1.50)	(0.0,16.5)	(0.0,30.0)	(23.0,30.0)	(0.0,30.0)		(0.0,30.0)	(0.0,2.5)	(0.0,30.0)
BondRate	(0.64,1.44)		(0.0,13.0)						(0.0,28.0)	
Real BondRate	(-0.12,0.66)	(-0.50,-0.16)	(20.5,30.0)		(0.0,30.0)					
Earnings/Price	(0.28,0.90)	(0.34,0.86) (1.50,1.50)	(6.5,30.0)		(12.5,30.0)	(4.0,30.0)				
Div/Price	(0.50,1.26)	(-0.50,1.50)	(0.0,17.0)	(0.0,30.0)	(18.0,30.0)	(0.0,30.0)		(0.0,30.0)	(0.0,11.5)	(0.0,30.0)
Daily Time Series										
Abs.Returns	(0.08,0.88)	(1.38,1.50)	(5.0,30.0)		(3.0,30.0)			(0.5,4.0)		

Notes: The entries in the table show 95% confidence sets for the model parameters constructed by inverting the *S* and *H* tests. The sets were constructed by carrying the tests for a grid of values of parameter values, where $-0.49 \le d \le 1.49$, $0 \le c \le 30$, and $0 \le g \le 30$. The confidence sets are shown as intervals (a,b), where disconnected sets include more than one interval.

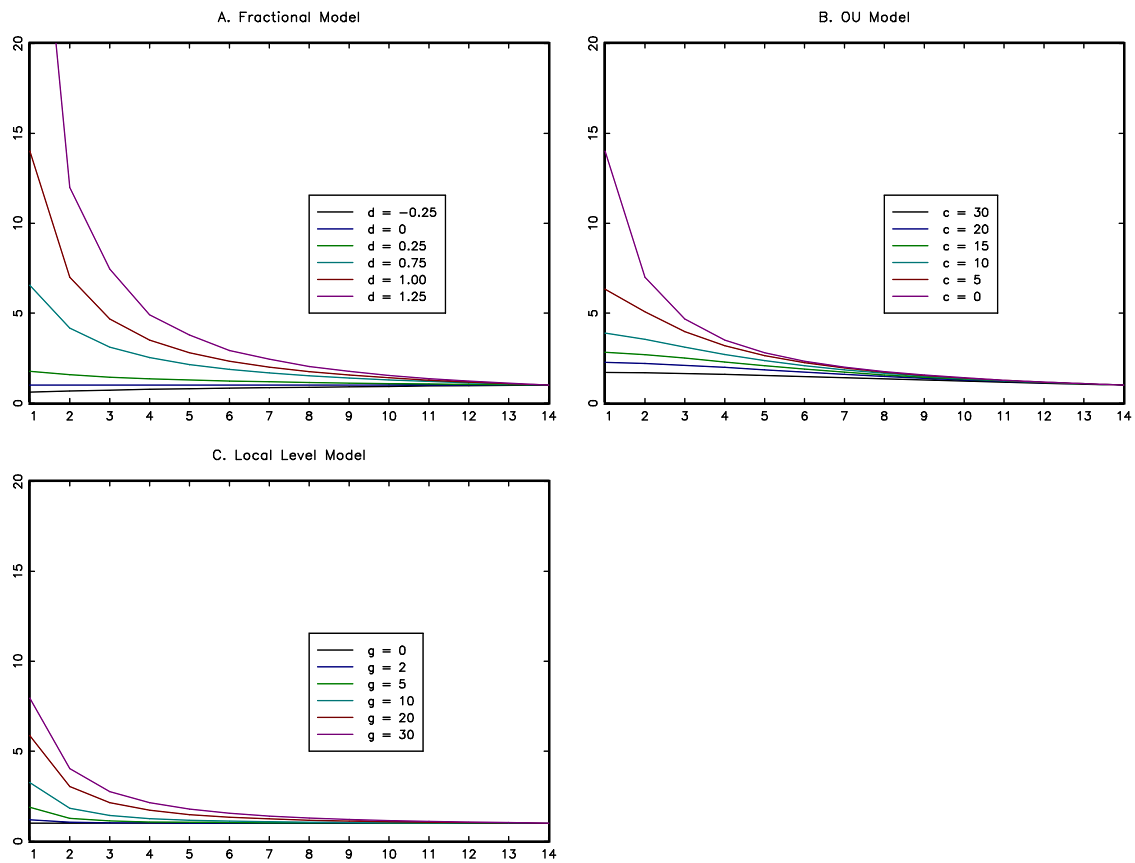
Figure 1
R^2 regression of $\sin(\pi\vartheta s+\phi)$ onto $\Psi_1(s)$ … $\Psi_q(s)$

A. Demeaned **B. Detrended**



Notes**:** These figures show the R^2 of a continuous time regression of a generic periodic series $\sin(\pi\vartheta s+\phi)$ onto the demeaned: column (a) or detrended: column (b) weight functions $\Psi_1(s)$, … , $\Psi_q(s)$, with q chosen such that $\Psi_j(s)$, j=1, … , q has frequency smaller or equal to ϑ_0=14. Panels (i) shows the R^2 value averaged over values of $\phi \in [0, \pi)$, panels (ii) shows the R^2 maximized over these values of ϕ for each ϑ, and panel (iii) shows the R^2 minimized over these values of ϕ for each ϑ. The solid curve in the first column (labeled Demeaned) shows results using the eigenfunctions $\varphi_l^\mu(s)$ from Theorem 1, and the dashed curve shows results using Fourier expansions. The solid curve in the second column (labeled Detrended) shows results using the eigenfunctions $\varphi_l^\tau(s)$ from Theorem 1, and the dashed curve shows results using detrended Fourier expansions.

Figure 2
Standard Deviation of X_l in Different Models



Notes: These figures show the square roots of the diagonal elements of $\Sigma_i(\theta)$ for different values of the parameter $\theta = (d, c, g)$, i denotes the covariance matrix for the fractional (panel a), *OU* (panel b), and *LL* models (panel c), computed using φ^{μ}. Larger values of d and g, and smaller values of c yield relatively larger standard deviations of X_1.

Figure 3
Asymptotic Rejection Frequencies for 5%-level LFST and LFUR Tests
(q =13, demeaned data)

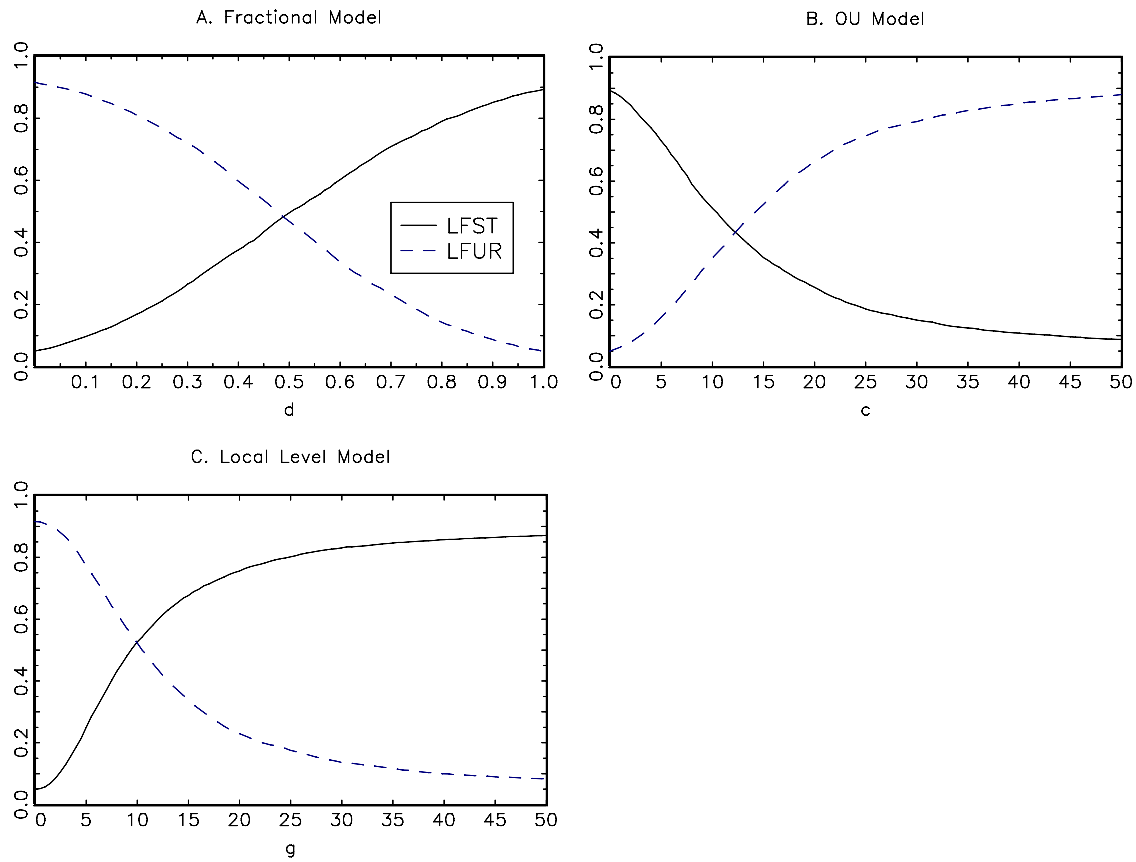
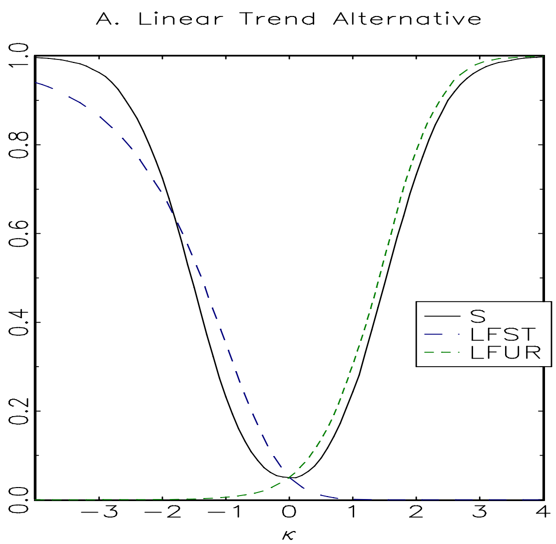
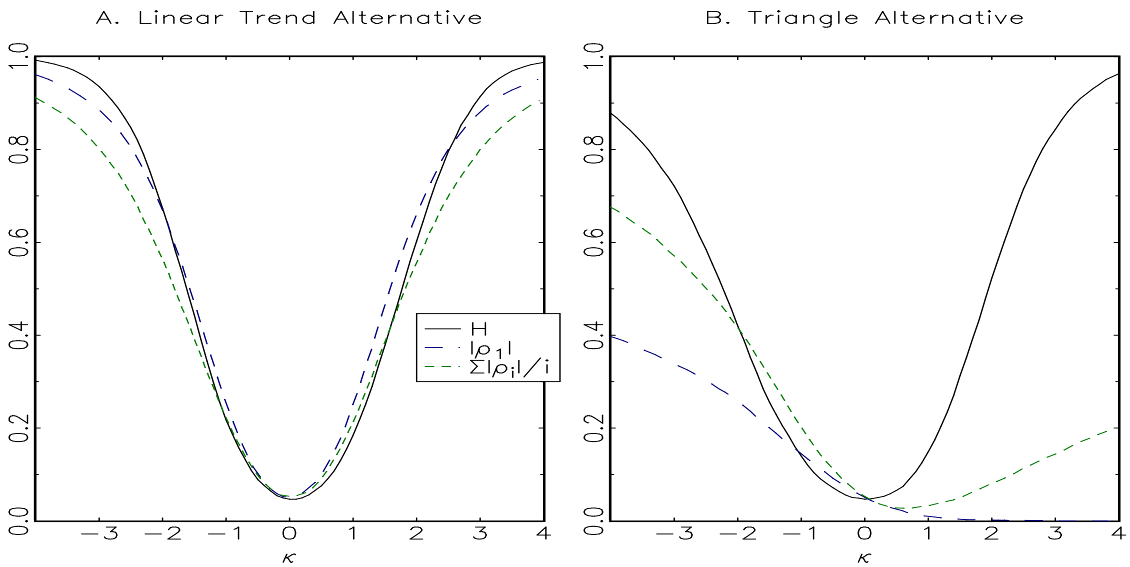


Figure 4
Power of 5%-Level *S*, LFST and LFUR Tests
(*q* =13, demeaned data)



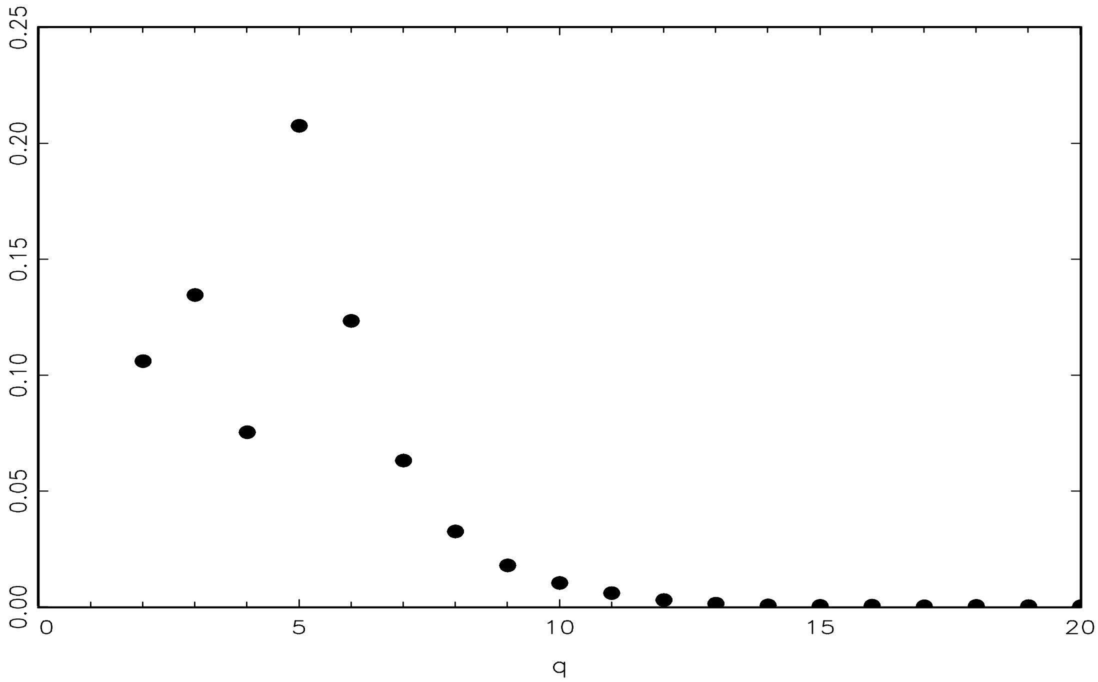
Notes: The alternatives have the form $\Sigma_1 = \Lambda\Sigma_0\Lambda$, where Λ= diag[exp(δ_1), … , exp(δ_{13})], where $\delta_i = [(i–1)/(q–1)]\times\kappa$ in panel (a), and $\delta_i = [(i–1)/6]\times\kappa$ for $i \leq 7$ and $\delta_i = \delta_{14-i}$ for $8 \leq i \leq 13$ in panel (b). The LFST results are for the I(0) model for Σ_0, the LFUR results are for the *I*(1) model, and the *S* results are for a model with diagonal Σ_0.

Figure 5
Power of 5%-Level H, $|\rho_1|$, and $\sum_{i=1}^{q-1} |\rho_i|/i$ Tests for the $I(0)$ model
(q =13, demeaned data)



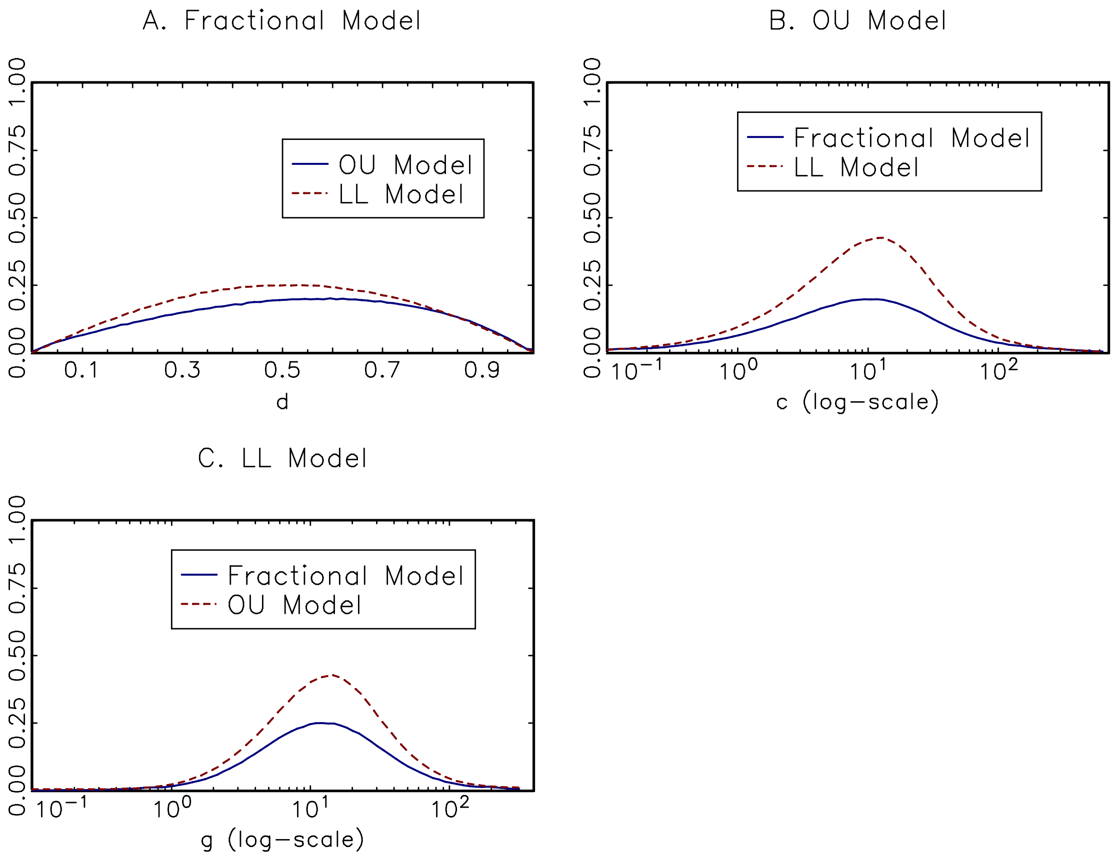
Notes: The alternatives are generated by the $I(0)$ model with $\ln[h(s)] = s\kappa$ in panel (a), and $\ln[h(s)] = s\kappa$ for $s \leq ½$ and $\ln[h(s)] = \kappa(1 - s)$ for $s > ½$ in panel (b).

Figure 6
p*-values for the LFST test as a function of *q



Notes: The figure shows the *p*-value for LFST computed using (X_{1T}, … , X_{qT}) constructed from demeaned values of the logarithm of the real $/£ exchange rate using annual observations from 1791-2004. The cutoff period (in years) corresponding to q can be computed as $period = 2T/q$, with $T = 214$.

Figure 7
Total Variation Distance



Notes: Results are shown for the demeaned case with q = 13.