

Long-Run Covariability*

Ulrich K. Müller and Mark W. Watson

Princeton University

Department of Economics

Princeton, NJ, 08544

First draft: August 2016

This version: December 2017

Abstract

We develop inference methods about long-run comovement of two time series. The parameters of interest are defined in terms of population second-moments of low-frequency transformations (“low-pass” filtered versions) of the data. We numerically determine confidence sets that control coverage over a wide range of potential bivariate persistence patterns, which include arbitrary linear combinations of $I(0)$, $I(1)$, near unit roots and fractionally integrated processes. In an application to U.S. economic data, we quantify the long-run covariability of a variety of series, such as those giving rise to balanced growth, nominal exchange rates and relative nominal prices, the unemployment rate and inflation, money growth and inflation, earnings and stock prices, etc.

JEL classification: C22, C53, E17

Keywords: band-pass regression, fractional integration, great ratios

*This work was presented as the Fisher-Schultz lecture at the 2016 European meetings of the Econometric Society. We thank David Papell, Ken West, three referees and participants at several seminars for their comments. Support was provided by the National Science Foundation through grant SES-1627660.

1 Introduction

Economic theories often have stark predictions about the covariability of variables over long-horizons: consumption and income move proportionally (permanent income/life cycle model of consumption) as do nominal exchange rates and relative nominal prices (long-run PPP), the unemployment rate is unaffected by the rate of price inflation (vertical long-run Phillips curve), and so forth. But there is a limited set of statistical tools to investigate the validity of these long-run propositions. This paper expands this set of tools.

Two fundamental problems plague statistical inference about long-run phenomena. The first is the paucity of sample information: there are few “long-run” observations in the samples typically used in empirical analyses of long-run relations. The second is that inference critically depends on the data’s long-run persistence. Random walks yield statistics with different probability distributions than i.i.d. data, for example, and observations from persistent autoregressions or fractionally integrated processes yield statistics with their own unique probability distributions. Taken together these two problems conspire to make long-run inference particularly difficult: proper inference depends critically on the exact form of long-run persistence, but there is limited sample information available to empirically determine this form.

This paper develops methods designed to provide reliable inference about long-run covariability for a wide range of persistence patterns. The methods rely on a relatively small number of low-frequency averages of the data to measure the data’s long-run variability and covariability. Our focus is on parameters that characterize the population second moments of these low-frequency averages.

Our main contribution is to provide empirical researchers with a relatively easy-to-use method for constructing confidence intervals for long-run correlation coefficients, linear regression coefficients, and standard deviations of regression errors. These confidence intervals are valid for $I(0)$, $I(1)$, near unit roots, fractionally integrated models, and linear combinations of variables with these forms of persistence. Using a set of pre-computed “approximate least favorable distributions,” the confidence intervals readily follow from the formulae discussed in Section 4.¹

The outline of the paper is as follows. The next section introduces two empirical examples,

¹The replication files contains a matlab function for computing these confidence intervals, available at www.princeton.edu/~mwatson.

the long-run relationship between consumption and GDP and between short- and long-term nominal interest rates, and defines the notion of long-run variability and covariability used throughout the paper. Our definition involves the population second moments of long-run projections, where these projections are similar to low-pass filtered versions of the data (e.g., Baxter and King (1999) and Hodrick and Prescott (1997)). In the long-run projections we employ, long-run covariability is equivalently captured by the covariability of a small number q of trigonometrically weighted averages of the data. The population second moments of the projections therefore correspond to an average of the spectrum (or pseudo-spectrum when the spectrum does not exist) over a narrow low-frequency band. Thus, this paper's long-run covariance parameters are those from low-frequency band spectrum regression (as in Engle (1974)), extended to allow for processes with more than $I(0)$ persistence. A key distinction between this paper and previous semiparametric approaches to the joint low-frequency behavior of persistent time series (see, for instance, Phillips (1991), Marinucci and Robinson (2001), Chen and Hurvich (2003), Robinson and Hualde (2003), Robinson (2008) or Shimotsu (2012)) is that in our asymptotic analysis, we keep q fixed as a function of the sample size. This ensures that the small sample paucity of low-frequency information is reflected in our asymptotic approximations, as in Müller and Watson (2008), which yields more reliable inference in samples typically used in empirical macroeconomics.

Section 3 derives the large-sample normality of the q pairs of trigonometrically weighted averages and introduces a flexible parameterization of the joint long-run persistence properties of the underlying stochastic process. The large-sample framework developed in Section 3 reduces the problem of inference about long-run covariability parameters into the problem of inference about the covariance matrix of a $2q$ dimensional multivariate normal random vector. Section 4 reviews relevant methods for solving this parametric small sample problem. Section 5 uses the resulting inference methods and post-WWII data from the United States to empirically study several familiar long-run relations involving balanced growth (GDP, consumption, investment, labor income, and productivity), the term structure of interest rates, the Fisher correlation (inflation and interest rates), the Phillips correlation (inflation and unemployment), PPP (exchange rates and price ratios), money growth and inflation, consumption growth and real returns, and the long-run relationship between stock prices, dividends and earnings.

2 Long-run projections and covariability

2.1 Two empirical examples of long-run covariability

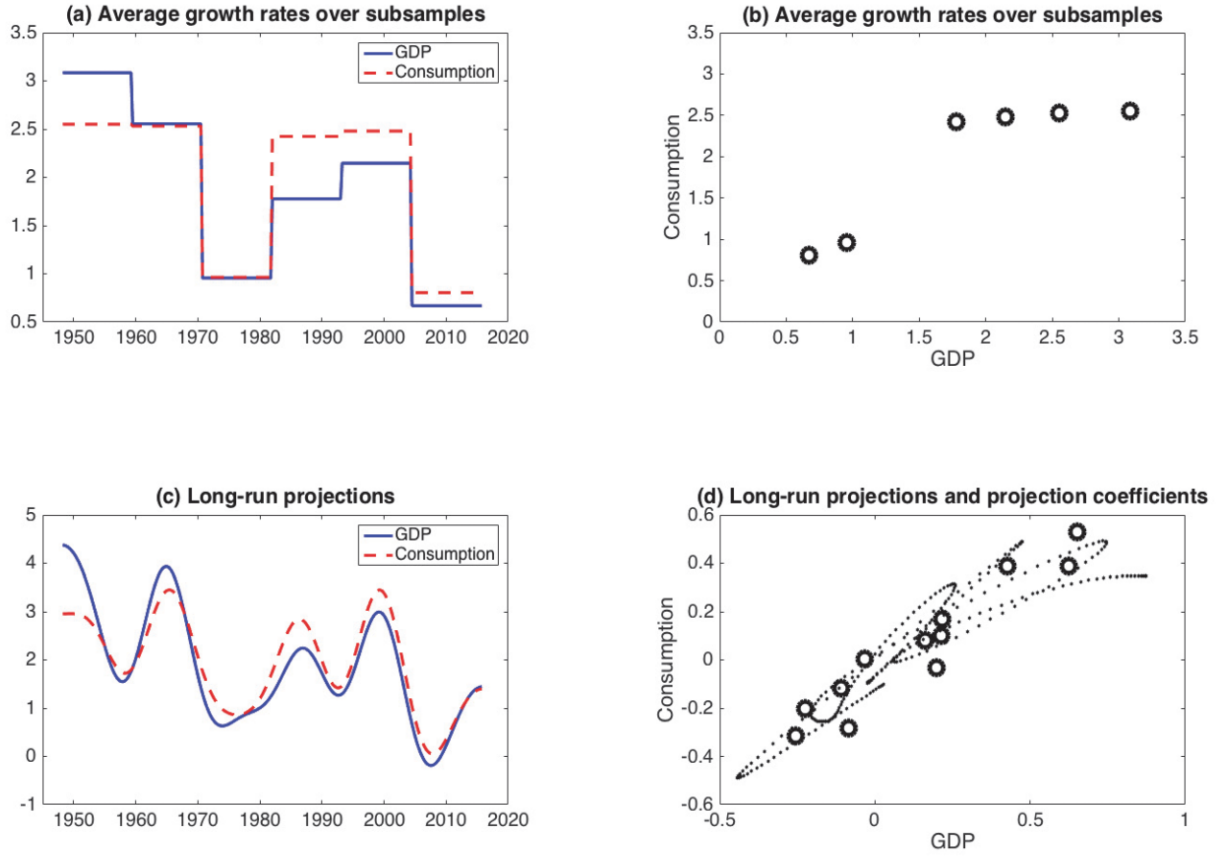
We begin by examining the long-run covariability of GDP and consumption and of short- and long-term nominal interest rates. These data motivate and illustrate the methods developed in this paper.

Consumption and income: One of the most celebrated and studied long-run relationship in economics concerns income and consumption. The long-run stability of the consumption/income ratio is one of economics’ “Great Ratios” (Klein and Kosobud (1961)) and even a casual glance at the U.S. data suggests the two variables move together closely in the long run. Consider, for example, the evolution of U.S. real per-capita GDP and consumption over the post-WWII period. In the 17 years from 1948 through 1964, GDP increased by 62% and consumption increased by 52%. Over the next 17 years (1965-1981) both GDP and consumption grew more slowly, by only 30%. Growth rebounded during 1982 to 1998, when GDP grew by 43% and consumption increased 55%, but slowed again over 1999-2015 when GDP grew by only 17% and consumption increased by only 23%. Over these 17-year periods, there was substantial variability in the average annual rate of growth of GDP (2.9%, 1.4%, 2.1%, and 0.9% per year, respectively over the sub-samples), and these changes were roughly matched by consumption (annual average growth rates of 2.5%, 1.5%, 2.6%, and 1.2%). Thus, over periods of 17 years, GDP and consumption exhibited substantial long-run variability and covariability in the post-WWII sample period.²

With this in mind, the first two panels of Figure 1 plot the average growth rates of GDP and consumption over six non-overlapping sub-samples in 1948-2015. Figure 1.a plots the average growth rates against time, and Figure 1.b is a scatterplot of the six average growth rates for consumption against the corresponding values for GDP. Each of the six sub-samples contains roughly 11 years, spans of history longer than the typical business cycle, and in this sense capture “long-run” variability in GDP and consumption. Average GDP and consumption growth over these subsamples exhibited substantial variability and (from

²Consumption is personal consumption expenditures (including durables) from the NIPA; Section 5 shows results for non-durables, services, and durables separately. Both GDP and consumption are deflated by the PCE deflator, so that output is measured in terms of consumption goods, and expressed in per-capita terms using the civilian non-institutionalized population over the age of 16. The supplemental appendix contains data sources and descriptions for all data used in this paper.

Figure 1: Long-run average growth rates of consumption and GDP



Notes: Panel (a) shows sample averages of the variables over the period shown. Panels (b) is a scatterplot of the variables in (a). Panel (c) plots the projections of the data onto the low-frequency cosine terms discussed in the text, where sample means have been added to projections so they are consistent with the averages plotted in Figure 1(a). The small dots in panel (d) are a scatterplot of the variables in (c) (after scaling) and the large circles are a scatter plot of the projection coefficients (X_{JT}, Y_{JT}) from (c).

the scatter plot) roughly one-for-one covariability.

Figure 1.c sharpens the analysis by plotting “low-pass” transformations of the series designed to isolate variation in the series with periods longer than 11 years, computed as projections onto low-frequency periodic functions. These low-frequency projections produce series that are essentially the same as low-pass moving averages,³ but are easier to analyze. These low-frequency projections are computed as follows. Let x_t , $t = 1, \dots, T$ denote a time series (e.g., growth rates of GDP or consumption). We use cosine functions for the periodic functions; let $\Psi_j(s) = \sqrt{2} \cos(js\pi)$ denote the function with period $2/j$ (where the factor $\sqrt{2}$ simplifies a calculation below), $\Psi(s) = [\Psi_1(s), \Psi_2(s), \dots, \Psi_q(s)]'$ denote a vector of these functions with periods 2 through $2/q$, and Ψ_T denote the $T \times q$ matrix with t 'th row given by $\Psi((t - 1/2)/T)'$, so the j 'th column of Ψ_T has period $2T/j$. The GDP and consumption data in Figure 1 span $T = 272$ quarters, so setting $q = 12$ captures periodicities longer than $272/6 \approx 45$ quarters, or 11.3 years. The projection of x_t onto $\Psi((t - 1/2)/T)$ for $t = 1, \dots, T$ yields the fitted values

$$\hat{x}_t = X_T' \Psi((t - 1/2)/T) \quad (1)$$

where X_T are the projection (linear regression) coefficients, $X_T = (\Psi_T' \Psi_T)^{-1} \Psi_T' x_{1:T}$, and $x_{1:T}$ is the $T \times 1$ vector with t 'th element given by x_t . The fitted values from these projections, say (\hat{x}_t, \hat{y}_t) , are plotted in Figure 1.c and evolve much like the averages in Figure 1.a, but better capture the smooth transition from high-growth to low-growth periods.

The matrix Ψ_T has two properties that simplify calculations and interpretation of the long-run projections. First, $\Psi_T' l_T = 0$ where l_T is a vector of ones, so that \hat{x}_t also corresponds to the projection of $x_t - \bar{x}_{1:T}$ onto $\Psi((t - 1/2)/T)$, where $\bar{x}_{1:T}$ is the sample mean. More generally, X_T and \hat{x}_t are invariant to location shifts in the x_t -process, so with $x_t = \mu + u_t$, the properties of X_T and \hat{x}_t do not depend on the typically unknown value of μ .⁴ Second, $T^{-1} \Psi_T' \Psi_T = I_q$, so X_T corresponds to simple cosine-weighted averages of the data (i.e., are the “cosine transforms” of $\{x_t\}$)

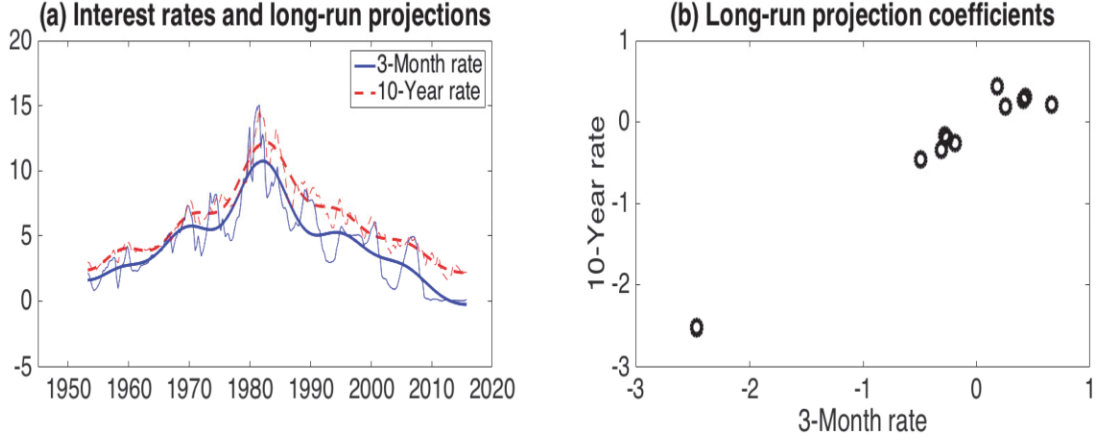
$$X_T = T^{-1} \Psi_T' x_{1:T}. \quad (2)$$

The orthogonality of the cosine regressors Ψ_T leads to a tight connection between the variability and covariability in the long-run projections (\hat{x}_t, \hat{y}_t) plotted in Figure 1.c and the

³The supplementary appendix provides a comparison.

⁴If the x_t -process contains a linear trend, say $x_t = \mu_0 + \mu_1 t + u_t$, then alternative periodic functions that are orthogonal to a time trend can be used so that X_T and \hat{x}_t do not depend on (μ_0, μ_1) . See Müller and Watson (2008) for one set of such functions.

Figure 2: Long-run projections of short- and long-term interest rates



Notes: Panel (a) plots the projections of the data onto the low-frequency cosine terms discussed in the text, where sample means have been added to projections so they are consistent with the low-pass moving averages plotted in Figure 1(c). Panel (b) is a scatter plot of the projection coefficients (X_{jT}, Y_{jT}) from (a).

cosine transforms (X_{jT}, Y_{jT}) :

$$T^{-1} \sum_{t=1}^T \begin{pmatrix} \hat{x}_t \\ \hat{y}_t \end{pmatrix} \begin{pmatrix} \hat{x}_t & \hat{y}_t \end{pmatrix} = T^{-1} \begin{pmatrix} X'_T \\ Y'_T \end{pmatrix} \Psi'_T \Psi_T \begin{pmatrix} X_T & Y_T \end{pmatrix} = \begin{pmatrix} X'_T X_T & X'_T Y_T \\ Y'_T X_T & Y'_T Y_T \end{pmatrix}. \quad (3)$$

Thus, the sample covariability of the T time series projections (\hat{x}_t, \hat{y}_t) coincides with the sample covariability of the q projection coefficients/cosine transforms (X_T, Y_T) .⁵ This is shown in Figure 1.d which shows a scatter plot of (scaled versions) of the projections (\hat{x}_t, \hat{y}_t) , shown as small dots, and the projection coefficients (X_{jT}, Y_{jT}) , shown as large circles. While the scatter plots capture the same variability and covariability in long-run movements in GDP and consumption growth, the projection coefficients eliminate much of the serial correlation evident in the (\hat{x}_t, \hat{y}_t) scatterplot.

Short-term and long-term interest rates. The second empirical example involves short- and long-term nominal interest rates, as measured by the rate on 3-month U.S. Treasury

⁵ Alternative low-frequency weights, such as Fourier transforms have the same orthogonality properties and could be used in place of the cosine transforms. While the general analysis accommodates these alternative weights, our numerical analysis uses the cosine weights presented in the text.

bills, x_t , and the rate on 10-year U.S. Treasury bonds, y_t , from 1953 through 2015. Figure 2 plots the levels of short- and long-term interest rates, (x_t, y_t) , along with their long-run projections, (\hat{x}_t, \hat{y}_t) , and cosine transforms, (X_T, Y_T) . This sample includes only 63 years, so periodicities longer than 11 years are extracted using long-run projections with $q = 11$. Figure 2.a shows that these long-run projections capture the rise in interest rates from the beginning of the sample through the early 1980s and then their subsequent decline. The projections for long-term interest rates closely track the projections for short-term rates and, given the connection between the projections and cosine transforms, X_{jT} and Y_{jT} are highly correlated (Figure 2.b).

These two datasets differ markedly in their persistence: GDP and consumption growth rates are often modeled as low-order MA models, while nominal interest rates are highly serially correlated. Yet, the variables in both data sets exhibit substantial long-run variation and covariation which is readily evident in the long-run projections (\hat{x}_t, \hat{y}_t) or equivalently (from (3)) the projection coefficients (X_T, Y_T) . This suggests that the covariance/variance properties of (X_T, Y_T) are a useful starting point for defining the long-run covariability properties of stochastic processes exhibiting a wide range of persistent patterns.

2.2 A measure of long-run covariability using long-run projections

A straightforward definition of long-run covariability is based on the population analogue of the sample second moment matrices in (3). Let Σ_T denote the covariance matrix of $(X'_T, Y'_T)'$, partitioned as $\Sigma_{XX,T}$, $\Sigma_{XY,T}$, etc., and define

$$\begin{aligned}\Omega_T &= T^{-1} \sum_{t=1}^T E \left[\begin{pmatrix} \hat{x}_t \\ \hat{y}_t \end{pmatrix} \begin{pmatrix} \hat{x}_t & \hat{y}_t \end{pmatrix} \right] \\ &= \sum_{j=1}^q E \left[\begin{pmatrix} X_{jT} \\ Y_{jT} \end{pmatrix} \begin{pmatrix} X_{jT} & Y_{jT} \end{pmatrix}' \right] = \begin{pmatrix} \text{tr}(\Sigma_{XX,T}) & \text{tr}(\Sigma_{XY,T}) \\ \text{tr}(\Sigma_{YX,T}) & \text{tr}(\Sigma_{YY,T}) \end{pmatrix}\end{aligned}\tag{4}$$

where the equalities directly follow from (3).

The 2×2 matrix Ω_T is the average covariance matrix of the long-run projections (\hat{x}_t, \hat{y}_t) in a sample of length T , and provides a summary of the variability and covariability of the long-run projections over repeated samples. Equivalently, by the second equality, Ω_T also measures the covariability of the cosine transforms (X_T, Y_T) . Corresponding long-run

correlation and linear regression parameters follow from the usual formulae

$$\begin{aligned}\rho_T &= \Omega_{xy,T} / \sqrt{\Omega_{xx,T} \Omega_{yy,T}} \\ \beta_T &= \Omega_{xy,T} / \Omega_{xx,T} \\ \sigma_{y|x,T}^2 &= \Omega_{yy,T} - (\Omega_{xy,T})^2 / \Omega_{xx,T}\end{aligned}\tag{5}$$

where $(\Omega_{xy,T}, \Omega_{xx,T}, \Omega_{yy,T})$ are the elements of Ω_T . The linear regression coefficient β_T solves the population least-squares problem

$$\beta_T = \underset{b}{\operatorname{argmin}} E \left[T^{-1} \sum_{t=1}^T (\hat{y}_t - b\hat{x}_t)^2 \right], \tag{6}$$

so that β_T is the coefficient in the population best linear prediction of the long-run projection \hat{y}_t by the long-run projection \hat{x}_t , $\sigma_{y|x,T}^2$ is the average variance of the prediction error, and ρ_T^2 is the corresponding population R^2 . These parameters thus measure the population linear dependence of the long-run variation of (x_t, y_t) . Equivalently, by the second equality in (4), β_T also solves

$$\beta_T = \underset{b}{\operatorname{argmin}} E \left[\sum_{j=1}^q (Y_{jT} - bX_{jT})^2 \right] \tag{7}$$

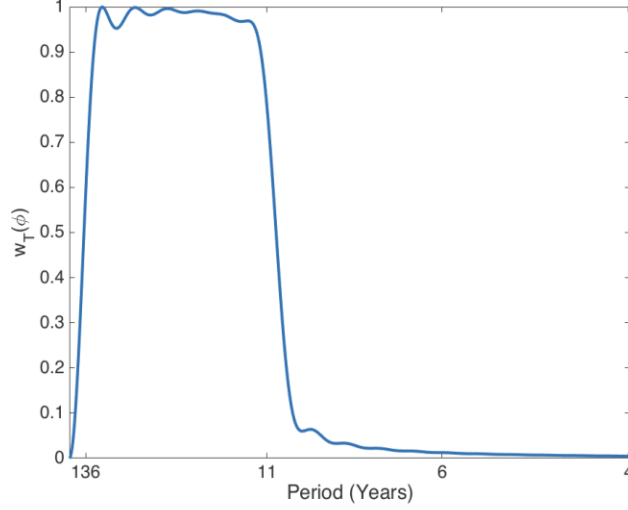
with a corresponding interpretation for $\sigma_{y|x,T}^2$ and ρ_T^2 . Thus, these parameters equivalently measure the (population) linear dependence in the scatter plots in Figures 2.c and 3.c.

The covariance matrix, Ω_T , or equivalently, $(\rho_T, \beta_T, \sigma_{y|x,T}^2)$, are the long-run population parameters that are the focus of our analysis. These parameters depend on the periods used to define the “long-run,” that is the value of q used to construct the long-run projections. In the empirical examples discussed above we chose periods longer than 11 years, that is periods longer than the U.S. business cycle. This led us to use $q = 12$ for GDP and consumption (with sample size $T = 272$ quarters) and $q = 11$ for interest rates (with sample size $T = 252$ quarters). If we had instead been interested in periods longer than 20 years we would have chosen $q = 7$ for GDP/consumption ($2T/q \approx 78$ quarters) and $q = 6$ for interest rates ($2T/q = 84$ quarters). The important point is that q defines the long-run periods of interest for the research question at hand.

2.3 Frequency domain interpretation of Ω_T and $(\rho_T, \beta_T, \sigma_{y|x,T}^2)$

The covariance matrix Ω_T has a natural and familiar frequency domain interpretation. Since (X_T, Y_T) are weighted averages of the data $z_t = (x_t, y_t)'$, Ω_T is a weighted av-

Figure 3: Spectral weight for low-frequencies with $T = 272$ quarters and $q = 12$



Notes: The figure plots $\left| \sum_{j=1}^q \sum_{t=1}^T \Psi_j((t-0.5)/T) e^{-it\phi} \right|^2$ for frequencies $0 \leq \phi \leq 2\pi/16$. The horizontal axis is labeled with periods per year ($= 0.5\pi/\phi$).

erage of the variances and covariances of z_t . If the time series are covariance stationary with spectral density matrix $F_z(\phi)$, these variances and covariances are weighted averages of the spectrum over different frequencies ϕ . In fact, a straightforward calculation shows (see the supplementary appendix) that $\Omega_T = (2\pi)^{-1} \int_{-\pi}^{\pi} F_z(\phi) w_T(\phi) d\phi$ where $w_T(\phi) = \left| \sum_{j=1}^q \sum_{t=1}^T \Psi_j((t-1/2)/T) e^{-it\phi} \right|^2$ and $i = \sqrt{-1}$.

Figure 3 plots the weights $w_T(\phi)$ for $T = 272$ quarters (the GDP-consumption sample size) and $q = 12$, where the horizontal axis shows periods (in years) instead of frequency (annual period $= 2\pi/(4\phi)$). The figure shows that Ω_T is essentially a bandpass version of the spectrum with periods between $2T$ (136 years) and $T/6$ (11.3 years) corresponding to cosine transforms with $j = 1$ through $j = 12$. Thus, Ω_T and the associated values of $(\rho_T, \beta_T, \sigma_{y|x,T}^2)$ are bandpass regression parameters, as in Engle (1974), for a particular low-frequency band. If, as in Engle's analysis, the data are generated by an $I(0)$ stochastic process, the spectrum is approximately flat over this band and inference follows relatively directly from classic results on spectral estimators (e.g., Brillinger (2001), Brockwell and Davis (1991)). However, if the process is not $I(0)$ in the sense that the spectrum (or pseudo-spectrum) is not flat over this

low-frequency band, $I(0)$ procedures lead to faulty inference akin to Granger and Newbold's (1974) spurious regressions. The goal of our analysis is to develop inference procedures that are robust to this $I(0)$ -flat-spectrum assumption.

As discussed above, the upper frequency cutoff $2T/q$, corresponding to 11 years in Figure 3, represents the highest low-frequency (shortest period) of interest for the researcher's analysis and is problem-specific. The lower cutoff $2T$, corresponding 136 years in Figure 3, is induced by the invariance to locations shifts of the cosine transforms. Without knowledge of the populations means, it is not possible to extract empirical information about arbitrarily low frequencies, and our estimand Ω_T reflects this impossibility. What is more, the fact that the weight $w_T(\phi)$ converges to zero as $\phi \rightarrow 0$ keeps our estimand Ω_T well-defined even for some (pseudo) spectra that diverge at frequency zero, such as for $I(1)$ processes.

3 Asymptotic approximations and parameterizing long-run persistence and covariability

The long-run correlation and regression parameters from Ω_T are functions of Σ_T , the covariance matrix of (X_T, Y_T) . This section takes up two related issues. The first is the asymptotic normality of the cosine-weighted averages (X_T, Y_T) . This serves as the basis for the inference methods developed in Section 4 and provides large-sample approximation for the matrices Σ_T and Ω_T . The second issue is a parameterization of long-run persistence and co-movement that determines the large-sample value of Σ_T and Ω_T .

3.1 Large-sample properties of long-run sample averages

Because (X_T, Y_T) are smooth averages of (x_t, y_t) , a central limit theorem effect suggests that these averages are approximately Gaussian random variables under a range of primitive conditions about (x_t, y_t) . The set of assumptions under which asymptotic normality holds turns out to be reasonably broad, and encompasses many forms of potential persistence. Specifically, with $z_t = (x_t, y_t)'$, suppose that Δz_t has moving average representation $\Delta z_t = C_T(L)\varepsilon_t$, where ε_t is a martingale difference sequence with non-singular covariance matrix, Δz_t has a spectral density $F_{\Delta z, T}$, and ε_t and $C_T(L)$ satisfy other moment and decay restrictions given in Müller and Watson (forthcoming, Theorem 1). The dependence of C_T and $F_{\Delta z, T}$ on the

sample size T accommodates many forms of persistence that require double arrays as data generating process, such as autoregressive roots of the order $1 - c/T$, for fixed c .⁶

If the spectral density converges for all frequencies close to zero

$$T^2 F_{\Delta z, T}(\omega/T) \rightarrow S_{\Delta z}(\omega) \quad (8)$$

in a suitable sense, then

$$\sqrt{T} \begin{pmatrix} X_T \\ Y_T \end{pmatrix} \Rightarrow \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \Sigma), \quad (9)$$

and the finite-sample second moment matrix correspondingly converges to its large-sample counterpart (Müller and Watson (forthcoming, Lemma 2))

$$T \text{Var} \begin{pmatrix} X_T \\ Y_T \end{pmatrix} = T \Sigma_T \rightarrow \Sigma. \quad (10)$$

The scaling necessary to achieve (8) is subsumed in $C_T(L)$. For instance, if x_t is a random walk, and y_t is i.i.d., then $C_T(L) = \text{diag}(1/T, 1 - L)$ induces (8).

The limiting covariance matrix Σ in (9) and (10) is a function of the (pseudo) “local-to-zero spectrum” $S_z(\omega) = S_{\Delta z}(\omega)/\omega^2$ of z_t and the cosine weights $\Psi_j(s)$ that determine (X_T, Y_T) ; specifically, from Müller and Watson (forthcoming)

$$\Sigma = \int_{-\infty}^{\infty} \left(I_2 \otimes \int_0^1 e^{i\lambda s} \Psi(s) ds \right)' S_z(\omega) \left(I_2 \otimes \int_0^1 e^{-i\lambda s} \Psi(s) ds \right) d\omega.$$

These limiting result are consistent with the results for $T = 272$ shown in Figure 3 in the previous section: (X_T, Y_T) are low-frequency weighted averages of the data and their covariance matrix depends of the (pseudo-) spectrum of (x_t, y_t) in a small band around frequency zero.

We make three comments about these large-sample results. First, they hold when the *first-difference* of z_t has a spectral density; the *level* of z_t is more persistent than its first difference and may have a (pseudo-) spectrum that diverges at frequency zero. In this case Σ remains finite because the cosine averages sum to zero ($\Psi_T' l_T = 0$), so they do not extract zero-frequency variation in the data. If the *level* of z_t has a spectral density then this restriction on the weights is not required and, for example, the centered sample mean of z_t

⁶We omit the corresponding dependence of $z_t = (x_t, y_t)$ on T to ease notation.

also has a large-sample normal limit. Second, for $z_t \sim I(d)$, the decay restrictions on $C_T(L)$ allow values of $d \in (-0.5, 1.5)$, which allows a reasonably wide range of persistent processes, but rules out some models of practical interest. For example, the first difference of an $I(0)$ process appended to a linear trend (i.e., the first difference of a “trend-stationary” process) is $I(-1)$ and is ruled out. And of course, it does not accommodate $I(d)$ processes with $d > 1.5$. Third, because $T\Sigma_T \rightarrow \Sigma$, also $T\Omega_T \rightarrow \Omega$ where Ω is defined as in the last expression of (4) with Σ in place of Σ_T . Correspondingly, $(\rho_T, \beta_T, T\sigma_{y|x,T}^2) \rightarrow (\rho, \beta, \sigma_{y|x}^2)$ with the limits defined by (5) with Ω in place of Ω_T . Thus, a solution to the problem of inference about $(\rho, \beta, \sigma_{y|x}^2)$ given observations (X, Y) readily translates into a solution to large-sample valid inference about $(\rho_T, \beta_T, \sigma_{y|x,T}^2)$ given (X_T, Y_T) , and, by invoking the arguments in Müller (2011), efficient inference in the former problem amounts to large sample efficient inference in the latter problem.

3.2 Parameterizing long-run persistence and covariability

The limiting average covariance matrix of the long-run projections, Ω , is a function of the covariance matrix of the cosine projections, Σ , which in turn is a function of the local-to-zero spectrum S_z of z_t . In this section we discuss parameterizations of S_z , and thus Σ and Ω .

It is constructive to consider two leading examples. In the first, z_t is $I(0)$ with long-run covariance matrix Λ . In this case, while the spectrum $F_{z,T}(\phi)$ potentially varies across frequencies $\phi \in [-\pi, \pi]$, the local-to-zero spectrum is flat $S_z(\omega) \propto \Lambda$. Straightforward calculations then show that $\Sigma = \Lambda \otimes I_q$ and $\Omega = \Lambda$, so the covariance matrix associated with the long-run projections corresponds to the usual long-run $I(0)$ covariance matrix. In this model, the cosine transforms (X_{jT}, Y_{jT}) plotted in Figures 1 and 2 are, in large samples and up to a deterministic scale, i.i.d. draws from a $\mathcal{N}(0, \Lambda)$ distribution. Inference about $\Omega = \Lambda$ and $(\rho, \beta, \sigma_{y|x}^2)$ thus follows from well-known small sample inference procedures for Gaussian data (see Müller and Watson (forthcoming)). In the second example, z_t is $I(1)$ with Λ the long-run covariance matrix for Δz_t . In this case $S_z(\omega) \propto \omega^{-2}\Lambda$, and a calculation shows that $\Sigma = \Lambda \otimes D$, where D is a $q \times q$ diagonal matrix with j 'th diagonal element $D_{jj} = (j\pi)^{-2}$. In this model, the cosine transforms (X_{jT}, Y_{jT}) plotted in Figures 1 and 2 are, in large samples and up to a deterministic scale, independent but heteroskedastic draws from $\mathcal{N}(0, (j\pi)^{-2}\Lambda)$ distributions. Thus $\Omega \propto \Lambda$, so the covariance matrix for long-run projections for z_t corresponds to the long-run covariance matrix for its first differences, Δz_t .

By weighted least squares logic, inference for $I(1)$ processes follows after reweighting the elements of (X_{jT}, Y_{jT}) by the square roots of the inverse of the diagonal elements of D and then using the same methods as in the $I(0)$ model.

GDP and consumption and short-term and long-term interest rates: Table 1 presents estimates and confidence sets for $(\rho_T, \beta_T, \sigma_{y|x,T})$ using (X_T, Y_T) for GDP and consumption (panel a) and for short- and long-term interest rates (panel b), where the focus is on periods longer than 11 years, so that $q = 12$ for panel (a) and $q = 11$ for panel (b). Results are presented for $I(0)$ and $I(1)$ models; a more general model of persistence is introduced below. The point estimates shown in the table are MLEs, and confidence intervals for $(\beta_T, \sigma_{y|x,T}^2)$ are computed using standard finite-sample normal linear regression formulae (after appropriate weighting in $I(1)$ model), and confidence sets for ρ_T are constructed as in Anderson (1984, section 4.2.2).

For GDP and consumption, there are only minor differences between the $I(0)$ and $I(1)$ estimates and confidence sets. The estimated long-run correlation is greater than 0.9, and the lower range of the 90% confidence interval exceeds 0.8 in both the $I(0)$ and $I(1)$ models. Thus, despite the limited long-run information in the sample (captured here by the 12 observations making up (X_T, Y_T)), the evidence points to a large long-run correlation between GDP and consumption. The long-run regression of consumption onto GDP yields a regression coefficient that is estimated to be 0.76 in the $I(0)$ model and 0.84 in $I(1)$ model. This estimate is sufficiently accurate that $\beta = 1$ is not included in the 90% $I(0)$ confidence set. The results for long-term and short-term nominal interest rates are similarly informative: there is strong evidence that the series are highly correlated over the long-run.

The validity of these confidence sets rests on the quality of the $I(0)$ and $I(1)$ models for approximating the spectral shape over the low-frequency band plotted in Figure 3. The $I(0)$ model assumes the spectrum is flat over this band, so the series behave like white noises for periods longer than 11 years, while the $I(1)$ model assumes random-walk behavior. Neither assumption is particularly compelling. Moreover, as we show in Table 4 below, the $I(0)$ assumption yields confidence intervals with coverage probability far below the nominal level when in fact the data were generated by the $I(1)$ model, and vice versa, and both yield faulty inference when persistence is something other than $I(0)$ or $I(1)$. With this motivation, the next subsection proposes a more flexible parameterization of persistence.

Table 1: Long-run covariability estimates and confidence intervals
using the $I(0)$ and $I(1)$ models: periods longer than 11 years

a. GDP and consumption

		ρ	β	$\sigma_{y x}$
$I(0)$	Estimate	0.93	0.76	0.36
	67% CI	0.87, 0.96	0.67, 0.85	0.30, 0.46
	90% CI	0.80, 0.97	0.60, 0.92	0.27, 0.55
$I(1)$	Estimate	0.93	0.84	0.35
	67% CI	0.88, 0.96	0.74, 0.94	0.29, 0.45
	90% CI	0.82, 0.97	0.66, 1.01	0.26, 0.54

b. Short- and long-term interest rates

		ρ	β	$\sigma_{y x}$
$I(0)$	Estimate	0.98	0.96	0.60
	67% CI	0.96, 0.98	0.90, 1.03	0.50, 0.79
	90% CI	0.93, 0.99	0.84, 1.08	0.44, 0.96
$I(1)$	Estimate	0.97	0.93	0.38
	67% CI	0.93, 0.98	0.85, 1.01	0.32, 0.50
	90% CI	0.90, 0.98	0.78, 1.07	0.28, 0.61

Notes: Periods longer than 11 years correspond to $q = 12$ for panel (a) and $q = 11$ for panel (b). The rows labeled "Estimate" are the maximum likelihood estimates using the large-sample distribution of the cosine transforms for the $I(0)$ and $I(1)$ models and the rows labeled CI denotes the associated confidence intervals.

3.2.1 (A, B, c, d) model

The shape of the local-to-zero spectrum determines the long-run persistence properties of the data, and misspecification of this persistence leads to faulty inference about long-run covariability. Simply put, reliable inference requires a parameterization of the spectrum that yields a good approximation to the persistence patterns of the variables over the low-frequency band under study. Addressing this issue faces a familiar trade-off: the parameterization needs to be sufficiently flexible to yield reliable inference about long-run covariability for a wide range of economically-relevant stochastic processes and yet be sufficiently parsimonious to allow meaningful inference with limited sample information. $I(0)$ persistence generates a flat local-to-zero spectrum, and $I(1)$ persistence generates a local-to-zero spectrum proportional to ω^{-2} . Both of these models are parsimonious, but tightly constrain the spectrum. This limits their usefulness as general models for conducting inference about long-run covariability.

With this trade-off in mind, we use a parameterization that nests and generalizes a range of models previously used to model persistence in economic time series. The parameterization is a bivariate extension of the univariate (b, c, d) model used in Müller and Watson (2016) and yields a local-to-zero spectrum of the form

$$S_z(\omega) \propto A \begin{pmatrix} (\omega^2 + c_1^2)^{-d_1} & 0 \\ 0 & (\omega^2 + c_2^2)^{-d_2} \end{pmatrix} A' + BB' \quad (11)$$

where A and B are 2×2 matrices with A unrestricted and B lower triangular.⁷

The primary motivation for this (A, B, c, d) model is as a parsimonious but flexible functional form for the local-to-zero spectrum. It combines and generalizes several standard spectral shapes. For example, with $A = 0$ it is the $I(0)$ local-to-zero spectrum. When $B = 0$, $c = 0$ and $d_1 = d_2 = 1$ it yields the $I(1)$ spectrum; $B = 0$, $d_1 = d_2 = 1$ yields a spectrum for arbitrary linear combinations of two independent local-to-unity processes with mean reversion parameters equal to c_1 and c_2 ; $B = 0$ and $c = 0$ yields a bivariate fractional spectrum with parameters d_1 and d_2 . Other choices of (A, B, c, d) yield spectra from models that combine persistent and non-persistent components (as in cointegrated or “local-level”

⁷This is the spectrum of a bivariate Whittle-Matérn (c.f., Lindgren (2013)) process with time series representation $z_t = A\tau_t + e_t$, where $\tau_t = (\tau_{1t}, \tau_{2t})'$ is a bivariate process with uncorrelated $\{\tau_{1t}\}$ and $\{\tau_{2t}\}$, $(1 - \phi_{i,T}L)^{d_i}\tau_{it} = T^{-d_i/2}\varepsilon_{it}$, $\phi_{i,T} = 1 - c_i/T$, $\varepsilon_t \sim I(0)$ with long-run variance equal to I_2 , $e_t \sim I(0)$ with long-run variance equal to BB' , and zero long-run covariance with ε_t .

models) but go beyond the usual $I(0)/I(1)$ or fractional formulations.

The nesting of the cointegrated model in the (A, B, c, d) model is particularly interesting because it too focuses on long-run relationship between the variables. Formally of course, cointegration concerns common patterns of *persistence* in the variables not their *variance* and *covariance*: in its canonical form x_t and y_t are cointegrated if both are $I(1)$ and yet a linear combination of the variables is $I(0)$. This implies that the variables share a single $I(1)$ trend in addition to $I(0)$ components. When the innovations of the $I(1)$ and $I(0)$ components are of the same order of magnitude, then the spectrum is dominated by the $I(1)$ component over low-frequencies. Thus, in the context of the (A, B, c, d) model, the canonical cointegration model corresponds to the restrictions $B = 0$ (because the marginal processes are dominated by the stochastic trends), $c_1 = c_2 = 0$ and $d_1 = d_2 = 1$ (so the trend components are $I(1)$), and A has rank 1 (because there is a single common trend). The singularity in A in turn induces singularities in S_z and Σ . The large- T limit of the usual formulation of cointegration thus implies that long-run projections computed with a fixed value of q are perfectly correlated, and the scatter plot of projection coefficients lie on straight line with slope corresponding to the cointegrating coefficient. Of course this perfect correlation will not obtain with finite T , so practically useful approximations would require a non-singular value of A and/or a non-zero value of B .

3.2.2 Beyond the (A, B, c, d) model

While the (A, B, c, d) parameterization encompasses many standard models, it is useful to highlight some models that are *not* encompassed by (11). We discuss two here.

One restriction of the (A, B, c, d) model is the asymptotic independence of the persistent and $I(0)$ components, as captured by A and B , respectively. One model where this independence is restrictive is in a cointegrated model with a local-to-zero cointegration coefficient and correlated $I(0)$ and trend components. Consider, for instance, a model where x_t is the stochastic trend, and y_t is a linear combination of an $I(0)$ error correction term and x_t , with a coefficient on x_t that is of order $1/T$.⁸ Allowing x_t to follow a local-to-unity process, the model is

$$\begin{aligned} x_t &= (1 - c/T)x_{t-1} + u_{x,t} \\ y_t &= \frac{\gamma}{T}x_t + u_{y,t} \end{aligned} \tag{12}$$

⁸Given the invariance restrictions we impose in subsection 4.2.1 below, such a transformation of the cointegration model is without loss of generality for inference about β_T and $\sigma_{y|x,T}^2$.

where $u_t = (u_{x,t}, u_{y,t})$ is $I(0)$ with long-run covariance Σ_u and elements σ_x^2 , σ_y^2 and $\lambda\sigma_x\sigma_y$ in obvious notation. The corresponding local-to-zero spectral density is

$$S_z(\omega) \propto \begin{pmatrix} \sigma_x & 0 \\ \gamma & \sigma_y \end{pmatrix} \begin{pmatrix} (\omega^2 + c^2)^{-1} & \lambda(c + i\omega)^{-1} \\ \lambda(c - i\omega)^{-1} & 1 \end{pmatrix} \begin{pmatrix} \sigma_x & \gamma \\ 0 & \sigma_y \end{pmatrix}. \quad (13)$$

This “local cointegration” spectral density is outside the (A, B, c, d) model whenever $\lambda \neq 0$. The complex-valued local-to-zero spectral density indicates the presence of lead and lag relationships that span a non-trivial fraction of the sample size. Since the level of x_t depends in a non-negligible manner on values of $u_{x,s}$ with $s \ll t$, x_t is correlated with values of $u_{y,s}$ in the distant past when $\lambda \neq 0$.

In this model, a calculation shows that the population regression of Y onto X has a regression coefficient $\beta = \gamma + \lambda c$, so it depends on the local cointegrating coefficient γ , the $I(0)$ correlation coefficient λ , and the persistence parameter c . In contrast, previous analyses of the model (12) (see, for instance, Elliott (1998) or Jansson and Moreira (2006)), focused on the cointegration parameter γ . This difference emphasizes the distinct goals of the analyses: The cointegration parameter γ yields the linear combination of (x, y) with minimum *persistence*, while the regression parameter β yields the linear combination with minimum *variance*. In general, there is no reason to expect that these are the same, and this makes it surprising that these two goals yield the same estimand ($\beta = \gamma$) in the canonical cointegration model with $c = 0$, even for $\lambda \neq 0$.

There are many other ways of modelling the joint long-run properties of z_t . As already mentioned, since the (A, B, c, d) model has a real valued local-to-zero spectrum, it rules out long-span leads and lags between the series. One simple way to generate such lags is via

$$\begin{aligned} x_t &= A_{11}u_{1,t} + A_{12}u_{2,t-\lfloor\delta T\rfloor} \\ y_t &= A_{21}u_{1,t} + A_{22}u_{2,t-\lfloor\delta T\rfloor} \end{aligned} \quad (14)$$

where $u_t \sim I(0)$ as above. The parameter $\delta \in \mathbb{R}$ measures the lead of $u_{1,t}$ relative to $u_{2,t-\lfloor\delta T\rfloor}$ as a fraction of the sample size. In this “ $I(0)$ long-lag” model the local-to-zero spectral density satisfies

$$S_z(\omega) \propto A \begin{pmatrix} \sigma_x & \lambda\sigma_x\sigma_y \exp(i\delta\omega) \\ \lambda\sigma_x\sigma_y \exp(-i\delta\omega) & \sigma_y \end{pmatrix} A' \quad (15)$$

with $[A] = A_{ij}$. When $\lambda \neq 0$, the long lags relating x and y yield a complex local-to-zero cross spectrum which is outside the (A, B, c, d) class.

Our construction does not guarantee that confidence sets have their desired coverage probabilities outside the (A, B, c, d) model; we present numerical results in Section 4.3 below that investigate the magnitude of inference errors associated with models (13) and (15).

4 Constructing confidence intervals for ρ , β , and $\sigma_{y|x}$

4.1 An overview

There are several approaches one might take to construct confidence intervals for the parameters ρ, β , and $\sigma_{y|x}$ from the observations (X, Y) . As a general matter, the goal is to compute confidence intervals that are as informative (“short”) as possible, subject to the coverage constraint that they contain the true value of the parameter of interest with a pre-specified probability. We construct confidence intervals by explicitly solving a version of this problem.

Generically, let θ denote the vector of parameters characterizing the probability distribution of (X, Y) , and let Θ denote the parameter space. (In our context, θ denotes the (A, B, c, d) -parameters.) Let $\gamma = g(\theta)$ denote the parameter of interest. ($\gamma = \rho, \beta$, or $\sigma_{y|x}$ for the problem we consider). Let $H(X, Y)$ denote a confidence interval for γ and $\text{lgth}(H(X, Y))$ denote the length of the interval. The objective is to choose H so it has small expected length, $E[\text{lgth}(H(X, Y))]$, subject to coverage, $P(\gamma \in H(X, Y)) \geq 1 - \alpha$, where α is a pre-specified constant. Because the probability distribution of (X, Y) depends on θ , so will the expected length of $H(X, Y)$ and the coverage probability. By definition, the coverage constraint must be satisfied for all values of $\theta \in \Theta$, but one has freedom in choosing the value of θ over which expected length is to be minimized. Thus, let W denote a distribution that puts weight on different values of θ , so the problem becomes

$$\min_H \int E_\theta(\text{lgth}(H(X, Y))) dW(\theta) \quad (16)$$

subject to

$$\inf_{\theta \in \Theta} P_\theta(\gamma \in H(X, Y)) \geq 1 - \alpha \quad (17)$$

where the objective function (16) emphasizes that the expected volume depends on the value of θ , with different values of θ weighted by W , and the coverage constraint (17) emphasizes that the constraint must hold for all values of θ in the parameter space Θ .

As noted by Pratt (1961), the expected length of confidence set for γ can be expressed in terms of the power of hypothesis tests of $H_0 : \gamma = \gamma_0$ versus $H_1 : \theta \sim W$. The solution to (16)-(17) thus amounts to the determination of a family of most powerful hypothesis tests, indexed by γ_0 . Elliott, Müller, and Watson (2015) suggest a numerical approach to compute corresponding approximate “least favorable distributions” for θ . We implement a version of those methods here; details are provided in the supplementary appendix. A key feature of the solution is that, conditional on the weighting function W and the least favorable distribution, the confidence sets have the familiar Neyman-Pearson form with a version of the likelihood ratio determining the values of γ included in the confidence interval.

While the resulting confidence intervals have smallest weighted expected length (up to the bounds used in the numerical approximation of the least favorable distributions), they can have unreasonable properties for particular realizations of (X, Y) . Indeed, for some values of (X, Y) , the confidence intervals might be empty, with the uncomfortable implication that, conditional on observing these values of (X, Y) , one is certain that the confidence interval excludes the true value. To avoid this, we follow Müller and Norets (2016) and restrict the confidence sets to be supersets of $1 - \alpha$ equal-tailed Bayes credible sets.

4.2 Some specifics

4.2.1 Invariance and equivariance

Correlations are invariant to the scale of the data. The linear regression of y_t onto x_t is the same as the regression of $y_t + bx_t$ onto x_t after subtracting b from the latter’s regression coefficient. It is sensible to impose the same invariance/equivariance on the confidence intervals. Thus, letting H^ρ , H^β , and H^σ denote confidence sets for ρ , β , and $\sigma_{y|x}$, we restrict these sets as follows:

$$\rho \in H^\rho(X, Y) \Leftrightarrow \rho \in H^\rho(b_x X, b_y Y) \text{ for } b_x b_y > 0 \quad (18)$$

$$\beta \in H^\beta(X, Y) \Leftrightarrow \frac{b_y \beta + b_{yx}}{b_x} \in H^\beta(b_x X, b_y Y + b_{yx} X) \text{ for } b_x, b_y \neq 0 \text{ and all values of } b_{yx} \quad (19)$$

$$\sigma_{y|x} \in H^\sigma(X, Y) \Leftrightarrow |b_y| \sigma_{y|x} \in H^\sigma(b_x X, b_y Y + b_{yx} X) \text{ for } b_x, b_y \neq 0 \text{ and all values of } b_{yx}. \quad (20)$$

These invariance/equivariance restrictions lead to two modifications to the solution to (16)-(17). First, they require the use of maximal invariants in place of the original (X, Y) .

The density of the maximal invariants for each of these transformations is derived in the supplementary appendix. Second, because the objective function (16) is stated in terms of (X, Y) , minimizing expected length by inverting tests based on the maximal invariant leads to a slightly different form of optimal test statistic. Müller and Norets (2016) develop these modifications in a general setting, and the supplementary appendix derives the resulting form of confidence sets for our problem.

4.2.2 Parameter space

We use the following parameter space for $\theta = (A, B, c, d)$: A and B are real, with B lower-triangular and (A, B) chosen so that Ω is non-singular, $c_i \geq 0$, and $-0.4 \leq d_i \leq 1$, for $i = 1, 2$.⁹ Thus, the confidence intervals control coverage over a wide range of persistence patterns including processes less persistent than $I(0)$, as persistent as $I(1)$, local-to-unity autoregressions, and where different linear combinations of x_t and y_t may have markedly different persistence. Even though our theoretical development would allow for values of d_i up to 1.5, we consider values of $d_i > 1$ reasonably rare in empirical analysis of economic time series. In order to obtain more informative inference, we therefore restrict the parameter space to $-0.4 \leq d_i \leq 1$, for $i = 1, 2$.

The confidence sets we construct require three distributions over θ : the weighting function W for computing the average length in the objective (16), the Bayes prior associated with the Bayes credible sets that serve as subsets for the confidence sets (Müller and Norets (2016)), and the least favorable distribution for θ that enforces the coverage constraint. The latter is endogenous to the program (16)-(17) and is approximated using numerical methods similar to those discussed in Elliott, Müller, and Watson (2015), with details provided in the supplementary appendix. In our baseline analysis we use the same distribution for W and the Bayes prior. Specifically, this distribution is based on the bivariate $I(d)$ model (so that $c_1 = c_2 = 0, B = 0$) with d_1 and d_2 independently distributed $U(-0.4, 1.0)$. Because of the invariance/equivariance restrictions, the scale of the matrix A is irrelevant and we set $A = R(\lambda_1)G(s)R(\lambda_2)$, where $R(\lambda)$ is a rotation matrix indexed by the angle λ , with λ_1 and λ_2 independently distributed $U[0, \pi]$. The relative eigenvalues of A are determined by the diagonal matrix $G(s)$, with $G_{11}/G_{22} = 15^s$ with s independently distributed $U[0, 1]$. We investigate the robustness of this choice in subsection 4.3 below.

⁹See the Appendix for additional details.

Table 2: Long-run covariability estimates, confidence intervals and credible sets using the (A, B, c, d) models: periods longer than 11 years

a. GDP and consumption				
		ρ	β	$\sigma_{y x}$
(A, B, c, d)	Estimate	0.91	0.77	0.41
	67% CI	0.83, 0.96	0.66, 0.87	0.33, 0.53
	90% CI	0.71, 0.97	0.48, 0.96	0.29, 0.66
	67% Bayes CS	0.83, 0.96	0.66, 0.87	0.33, 0.53
	90% Bayes CS	0.71, 0.97	0.58, 0.96	0.29, 0.66

b. Short- and long-term interest rates				
		0.96	0.95	0.63
(A, B, c, d)	Estimate	0.96	0.95	0.63
	67% CI	0.92, 0.98	0.87, 1.07	0.49, 0.97
	90% CI	0.89, 0.99	0.76, 1.16	0.42, 1.27
	67% Bayes CS	0.92, 0.98	0.87, 1.03	0.49, 0.82
	90% Bayes CS	0.89, 0.99	0.81, 1.09	0.42, 1.02

Notes: Periods longer than 11 years correspond to $q = 12$ for panel (a) and $q = 11$ for panel (b). The rows labeled "Estimate" are the posterior median based on the $I(d)$ model for the (A, B, c, d) model. "CI" denotes confidence interval, which is calculated as described in the text. "Bayes CS" are Bayes equal-tailed credible sets based on the posterior from the $I(d)$ model.

4.2.3 Empirical results for GDP, consumption, and interest rates

Table 2 shows estimates for $(\rho_T, \beta_T, \sigma_{y|x,T})$ and confidence sets using the (A, B, c, d) model. The estimated value of $(\rho_T, \beta_T, \sigma_{y|x,T})$ is the median of the posterior using the $I(d)$ -model prior, and the table also shows Bayes credible sets for this prior for comparison with the frequentist confidence intervals. For GDP and consumption, the (A, B, c, d) results look much like the results obtained for the $I(0)$ model shown in Table 1. For most entries, the Bayes credible sets are slightly larger than the $I(0)$ sets, presumably reflecting the possibility of persistence greater than $I(0)$. The frequentist confidence intervals often coincide with Bayes intervals, but occasionally are somewhat wider. The results indicate that GDP and consumption are highly correlated in the long-run (the 90% confidence set is $0.71 \leq \rho \leq 0.97$) and the long-run regression coefficient of consumption onto GDP is large, but less than unity (the 90% confidence set is $0.48 \leq \beta \leq 0.95$). The results for interest rates indicate that

Table 3: Asymptotic weighted average power of 10% level efficient test of $H_0: \rho = 0$ in (A, B, c, d) model ($q = 12$)

WAP computed for	WAP efficient test for						
	$W_{I(0)}$	$W_{I(1)}$	W_{base}		$W_{I(0)}$	$W_{I(1)}$	W_{base}
	without Bayes superset constraint				with Bayes superset constraint		
$W_{I(0)}$	0.69	0.34	0.69		0.64	0.34	0.64
$W_{I(1)}$	0.38	0.66	0.60		0.38	0.65	0.60
W_{base}	0.57	0.41	0.61		0.55	0.41	0.58

Notes: The entries are the asymptotic weighted average power over alternatives shown in rows using WAP efficient tests for alternatives given in columns.

long-run movements in short- and long-rates are highly correlated, and that these data are consistent with a unit long-run response of long-rates to short-rates.

4.3 Asymptotic power and size

In this subsection we investigate some aspects of power and size that govern the average length and coverage of the confidence intervals. For power we investigate the choice of the weighting function W and the restriction of our confidence sets to be supersets of $1 - \alpha$ Bayes credible sets. For size we investigate the non-coverage probability of confidence intervals constructed for misspecified models. To keep the discussion concise, we focus exclusively on tests of $H_0 : \rho = 0$ of level $\alpha = 10\%$ using $q = 12$ in this subsection.

We first consider variations in weighting functions W and the cost of imposing the Bayes superset restriction. We investigate three weighting functions: the baseline weighting function $W = W_{\text{base}}$ described above, the weighting function $W_{I(0)}$ that is equivalent to W_{base} except that $d_1 = d_2 = 0$, and $W_{I(1)}$ obtained by setting $d_1 = d_2 = 1$. For each W , we compute the weighted average power maximizing test, both with and without the restriction that the implied confidence set is a superset of the $1 - \alpha$ Bayes credible set. This results in six tests. For each test, we then compute its weighted average power, under each of the three weighting functions. In other words, we compute the power of the six tests of $H_0 : \rho = 0$ against the alternative that the data was generated by θ randomly drawn from W , for each of the three W . Table 3 shows the results. Consider first tests that do not impose the Bayes superset restriction. By construction, the test that is constructed to maximize weighted

Table 4: Asymptotic size of 10% level efficient test of $H_0: \rho = 0$ ($q = 12$)

Data generated by	Efficient test constructed for				
	$I(0)$	$I(1)$	$I(0)+I(1)$	$I(d)$	(A, B, c, d)
(A, B, c, d) models					
$I(0)$	0.10	1.00	0.10	0.09	0.09
$I(1)$	0.99	0.10	0.10	0.10	0.10
$I(0)+I(1)$	0.99	0.99	0.10	0.12	0.10
$I(d)$ with $-0.4 \leq d \leq 1.0$	0.99	1.00	0.31	0.10	0.10
(A, B, c, d)	0.99	1.00	0.31	0.13	0.10
non- (A, B, c, d) models					
Local cointegration	0.84	1.00	0.15	0.09	0.10
Long-lag $I(0)$	0.30	1.00	0.13	0.09	0.10
$I(d)$ with $-0.4 \leq d \leq 1.4$	1.00	1.00	0.46	0.42	0.47

Notes: Entries in the table are maximal null rejection frequencies for stochastic processes shown in rows for efficient tests shown in columns. The stochastic processes in the top panel are special cases of the (A, B, c, d) model with the parametric restrictions discussed in the paper. The stochastic processes in the bottom panel are not included in the paper's parameterization of the (A, B, c, d) model.

average power against a given W has the highest weighted average power against that W among all level α tests; these power-envelope values are shown in the diagonal entries in the table. The table indicates that the optimal test for $W_{I(0)}$ has substantially less power under $W_{I(1)}$ than this envelope, and vice versa. The test constructed under W_{base} , in contrast, is essentially on the envelope under $W_{I(0)}$, and loses only about 6 percentage points under $W_{I(1)}$. Turning to the comparison with tests that impose the Bayes superset restriction, the table suggests that its cost in terms of weighted average power is fairly small, especially under W_{base} .

We now turn to studying the size of tests in misspecified models. Results are shown in Table 4. The rows indicate the model used to generate the data, the columns show the model used to construct the test, and the entries are the null rejection probabilities, maximized over the parameters used to generate the data under the constraint of $\rho = 0$. We consider five tests that all maximize weighted average power against W_{base} , but that control size only in restricted versions of the (A, B, c, d) model by construction. None impose the Bayes superset constraint. The five models used to construct the tests are the the $I(0)$

model ($S_z(\omega) \propto BB'$), the $I(1)$ model ($S_z(\omega) = \omega^{-2}AA'$), a bivariate “local-level” that includes $I(0)$ and $I(1)$ components ($S_z(\omega) \propto \omega^{-2}AA' + BB'$), the fractional $I(d)$ model ($S_z(\omega) \propto ADA'$, D diagonal with $D_{jj} = \omega^{-2d_j}$) and the general (A, B, c, d) model with $S_z(\omega)$ given by (11), where for the last two, $-0.4 \leq d_1, d_2 \leq 1.0$, as in our baseline. We then compute the size of these tests under eight models. The first five are just the same as were used in the construction of the tests. The other three models are the local cointegration model (13), the long-lag $I(0)$ model (15) and the fractional model with $-0.4 \leq d_1, d_2 \leq 1.4$. The table therefore shows sizes computed from 40 experiments comprised of five different tests and data generated from eight different stochastic processes.

The top panel of the table shows results for data generated by each of the five models used to construct the tests, so the diagonal entries of the table are equal to 0.10 by construction. The off-diagonal entries larger than 0.10 indicate size distortions. For example, the 10-percent level $I(0)$ test mistakenly rejects for 99 percent of the draws when the data are generated by other models. The $I(1)$ test has similarly large size distortions when the data are not generated by the $I(1)$ model. These results mirror previous findings of the fragility of inference based on the assumption of exact $I(0)$ or $I(1)$ persistence patterns (e.g., den Haan and Levin (1997) for HAC inference in $I(0)$ models and Elliott (1998) for inference in cointegrated models). The $I(0) + I(1)$ model encompasses both the $I(0)$ and $I(1)$ models, so the associated test has good size control for these models, but has size equal to 31% in the $I(d)$ and (A, B, c, d) models. The $I(d)$ model encompasses the $I(0)$ and $I(1)$ models, and so controls size there by construction. It does not encompass the $I(0) + I(1)$ or (A, B, c, d) models, but exhibits only a relatively small size distortion in these cases.¹⁰ Since the models in the top panel are special cases of our baseline (A, B, c, d) model, the corresponding entries in the fifth column cannot be larger than 0.10 by construction.

The bottom panel of the table shows sizes for data generated by data outside our parameterization of the (A, B, c, d) model. The (A, B, c, d) model based inference seems robust to the long leads and lag patterns induced by the local cointegration model, and the $I(0)$ long lag model. Even though the Σ matrices induced by these models are quite distinct from those induced by the (A, B, c, d) model, this misspecification does not induce substantial overrejections. In contrast, allowing more persistence in the form of fractional stochastic trends with $d \geq 1.0$ can induce severe overrejections. Apparently, for purposes of inference

¹⁰Müller and Watson (2016) show that $I(d)$ model yields long-run prediction sets with significant under-coverage when data are generated by a univariate analogue of the (A, B, c, d) model, however.

about Ω , it is essential to allow for the correct persistence pattern of the marginals of z_t , while misspecifications of intertemporal dependence seems to play a lesser role.

5 Empirical Analysis

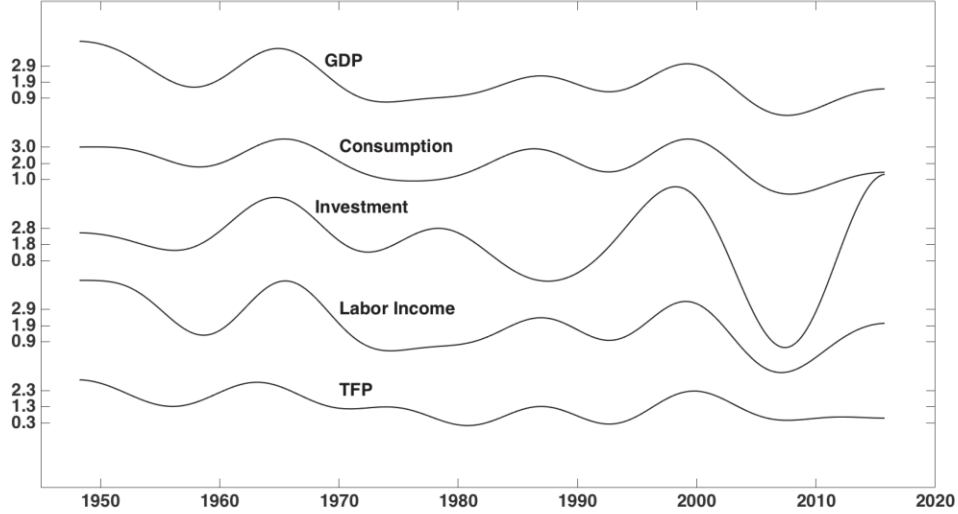
The last section showed results for the long-run covariation between GDP and consumption and between short- and long-term nominal interest rates. In this section we use the same methods to investigate other important long-run correlations. We focus on two questions: first, how much information does the sample contain about the long-run covariability, and second, what are the values of the long-run covariability parameters. A knee-jerk reaction to investigating long-run propositions in economics using, say, 68-year spans of data is that little can be learned, particularly so using analysis that is robust to a wide range of persistence patterns. In this case, even efficient methods for extracting relevant information from the data will yield confidence intervals that are so wide that they rule out few plausible parameter values. We find this to be true for some of the long-run relationships investigated below. But, as we have seen from the consumption-income and interest rate data, confidence intervals about long-run parameters can be narrow and informative, and this holds for several of the relationships that we now investigate.

Throughout the first two subsections we focus on periods longer than 11 years. For data available over the entire post-WWII period this entails setting $q = 12$. For shorter sample periods, smaller values of q are used, and these values are noted in context. The last subsection investigates the robustness of the empirical conclusions to focusing on periods longer than 20 years (so that $q = 6$ for data available over the entire post-WWII period).

5.1 Balanced growth correlations

In the standard one-sector growth model, variations in per-capita GDP, consumption, investment, and in real wages arise from variations in total factor productivity (TFP). Balanced growth means that the consumption-to-income ratio, the investment-to-income ratio, and labor's share of total income are constant over the long run. This implies perfect pairwise long-run correlations between the logarithms of income, consumption, investment, labor compensation, and TFP. In this model, the long-run regression of the logarithm of consumption onto the logarithm of income has a unit coefficient, as do the same regressions with

Figure 4: Long-run projections for GDP, consumption, investment, labor income and TFP growth rates: periods longer than 11 years



consumption replaced by investment or labor income. A long-run one-percentage point increase in TFP leads a long-run increase of $1/(1 - \alpha)$ percentage points in the other variables, where $(1 - \alpha)$ is labor's share of income. Of course, these implications involve the evolution of the variables over the untestable infinite long-run. That said, empirical analysis can determine how well these implications stand-up as approximations to below business cycle frequency variation in data spanning the post-WWII period. We use data for the U.S. and the methods discussed above to investigate these long-run balance growth propositions. The supplemental appendix contains a description of the data that are used.

Figure 4 plots the long-run projections of the growth rates of GDP, consumption, investment, labor income and TFP. (The long-run projections for consumption and GDP were shown previously in Figure 1.b.) The figure indicates substantial long-run covariability over the post-WWII period, but less so for investment than the other variables. Table 5 summarizes the results on the long-run correlations. The values above the main diagonal show point estimates constructed as the posterior median using the $I(d)$ -model with prior discussed above, together with 67% confidence intervals (shown in parentheses) using the general (A, B, c, d) model. The values below the main diagonal are the corresponding

Table 5: Long-run correlations of GDP, consumption, investment, labor compensation, and TFP: periods longer than 11 years

	GDP	Cons.	Inv.	$w \times n$	TFP
GDP		0.91 (0.83, 0.96)	0.53 (0.29, 0.72)	0.98 (0.96, 0.99)	0.78 (0.64, 0.89)
Cons.	(0.71, 0.97)		0.53 (0.30, 0.72)	0.90 (0.83, 0.96)	0.70 (0.49, 0.82)
Inv.	(0.02, 0.81)	(0.02, 0.81)		0.57 (0.34, 0.74)	0.38 (0.05, 0.60)
$w \times n$	(0.94, 0.99)	(0.55, 0.97)	(0.06, 0.82)		0.71 (0.53, 0.84)
TFP	(0.46, 0.95)	(0.29, 0.91)	(-0.08, 0.72)	(0.36, 0.93)	

Notes: All variables are measured in growth rates. The entries above the diagonal show the median of the posterior distribution followed by the 67% confidence interval. The entries below the diagonal show the 90% confidence interval.

90% confidence intervals using the (A, B, c, d) model. Table 6 reports results from selected long-run regressions.

As reported in the previous section, the long-run correlation between GDP and consumption is large. Labor income and GDP are highly correlated with a tightly concentrated 90% confidence interval of 0.94 to 0.99. The estimated long-run correlation of TFP and GDP is also high, although the correlation of TFP and the other variables appears to be somewhat lower. Investment and GDP are less highly correlated; the *upper* bound of the 90% confidence interval is only 0.8 and the lower bound is close to zero.

Table 6 shows results from long-run regressions of the growth rates of consumption, investment, and labor income onto the growth rate of GDP, and the corresponding regression of GDP onto TFP. Labor compensation appears to vary more than one-for-one with GDP and (as reported above) consumption less than one-for-one. The long-run investment-GDP regression coefficient is imprecisely estimated. Disaggregating consumption into nondurables, durables, and services, suggests that durable consumption responds more to long-run variations in GDP than do services and non-durables. These long-run regression results are reminiscent of results using business cycle covariability, and in Section 6 we investigate their robustness to the periodicities incorporated in the long-run analysis.

In summary, what has the 68-year post WWII sample been able to say about the balanced-growth implications of the simple growth model? First, that several of the variables are highly correlated over the long-run, defined as periods between 11 and 136 years, and second that the long-run regression coefficient on GDP is different from unity for some

Table 6: Selected long-run regressions involving GDP, consumption, investment, labor compensation, and TFP: periods longer than 11 years

Y	X		β			$\hat{\sigma}_{y x}$
			$\hat{\beta}$	67% CI	90% CI	
Consumption	GDP		0.77	0.66, 0.87	0.48, 0.96	0.41
Investment	GDP		1.23	0.68, 1.77	0.12, 2.24	2.19
Labor comp. ($w \times n$)	GDP		1.29	1.20, 1.36	1.14, 1.42	0.32
GDP	TFP		1.22	0.94, 1.49	0.72, 1.72	0.73
Cons. (Nondurable)	GDP		0.36	0.12, 0.59	-0.08, 0.76	0.89
Cons. (Services)	GDP		0.83	0.67, 0.99	0.54, 1.25	0.61
Cons. (Durables)	GDP		1.85	1.47, 2.26	1.19, 2.59	1.52
Inv. (Nonresidential)	GDP		0.97	0.39, 1.51	-0.05, 1.93	2.18
Inv. (Residential)	GDP		2.19	0.81, 3.57	-0.23, 4.69	5.64
Inv. (Equipment)	GDP		0.87	0.14, 1.55	-0.41, 2.12	2.77

Notes: All variables are measured in growth rates, in percentage points at an annual rate. The entries were constructed from the long-run regression of the variable labeled Y onto the variable labeled X .

variables (consumption and labor income). There is less information about the long-run covariability of investment with the other variables, although even here there are things to learn, such as the long-run correlation of investment and GDP is unlikely to be much larger than 0.8. Section 5.3 shows that similar results obtain using only periods longer than 20 years.

5.2 Other long-run relations

Figure 5 and Table 7 summarize long-run covariation results for an additional dozen pair of variables, using post-WWII U.S. data. (See the supplemental appendix for description and sources of the data.) We discuss each in turn.

CPI and PCE inflation. We begin with two widely-used measures of inflation, the first based on the consumer price index (CPI) and the second based on the price deflator for personal consumption expenditures (PCE). The Boskin Commission Report and related research (Boskin, Dulberger, Gordon, Griliches, and Jorgenson (1996), Gordon (2006)) highlights important methodological and quantitative differences in these two measures of inflation. For example, the CPI is a Laspeyres index, while the PCE deflator uses chain weighting, and this leads to greater substitution bias in the CPI. Differences in these inflation measures may change over time both because of the variance of relative prices (which affects substitution bias) and because measurement methods for both price indices evolved over the sample period.

Panel (a) of Figure 5 presents two plots; the first shows a time series plot of the long-run projections for PCE and CPI inflation, and the second shows the corresponding scatterplot of the projection coefficients, where the scatterplot symbols are the periods (in years) associated with the coefficients. For instance, the outlier “68.8” corresponds to the large negative coefficient on the first cosine function $\cos(\pi(t - 1/2)/T)$, which has a U-shape, and both inflation rates have a pronounced inverted U-shape in the sample. Long-run movements in PCE and CPI inflation track each other closely and the 90% confidence interval shown in Table 7 suggests that the long correlation is greater than 0.95. The long-run regression of CPI inflation on PCE inflation yields an estimated slope coefficient that is 1.13 (90% confidence interval: $0.98 \leq \beta \leq 1.24$) suggesting a larger bias in the CPI during periods of high trend inflation.

Long-run Fisher correlation and the real term structure: The next two entries in the

Figure 5: Long-run projections and projection coefficients: periods longer than 11 years

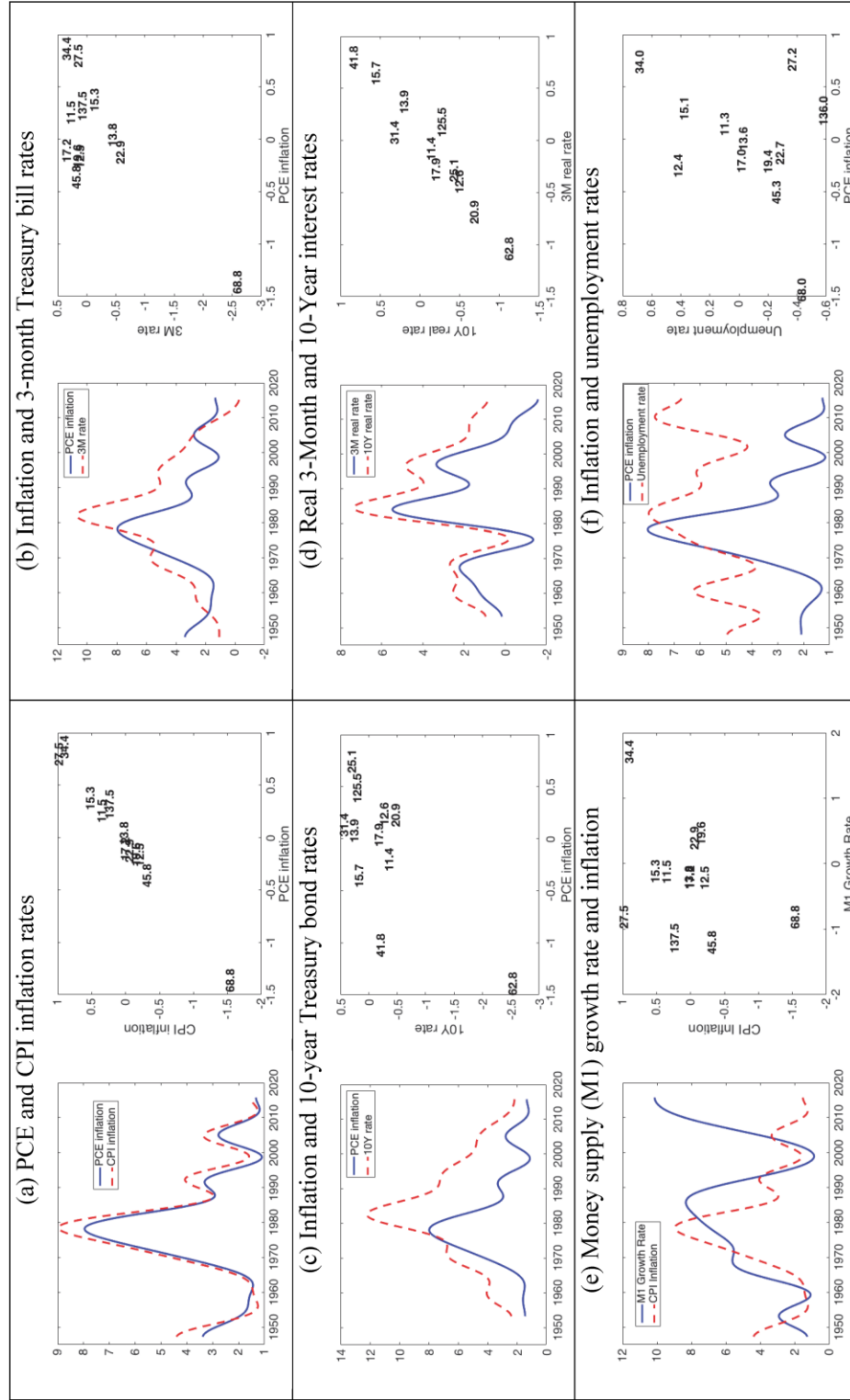
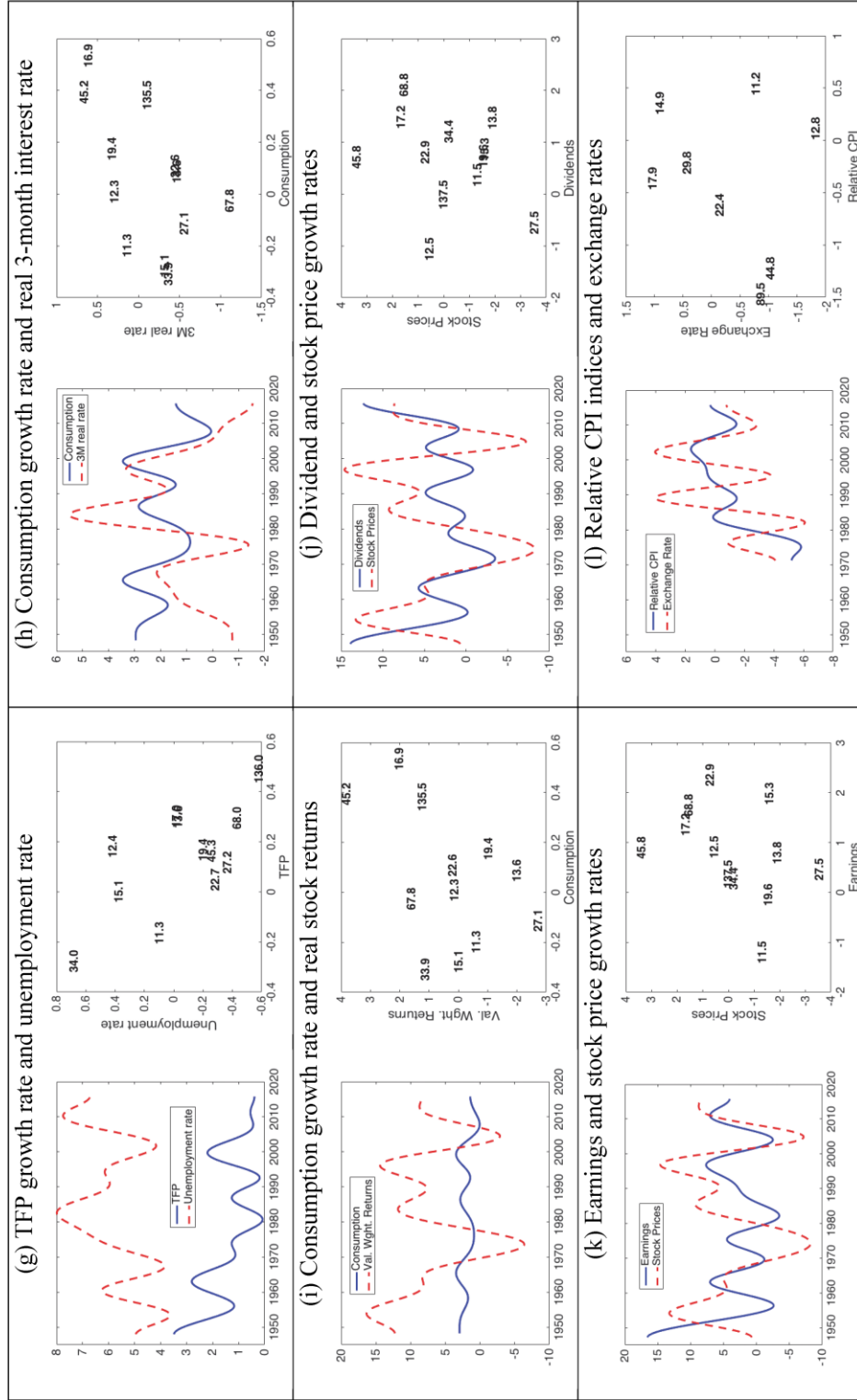


Figure 5 (continued)



Notes: The first plot in each panel shows the long-run projections of the time series. The second plot is a scatterplot of the long-run projection coefficients where the plot symbols indicate the period of the associated cosine function.

Table 7: Long-run covariation measures for selected variables: periods longer than 11 years

Y	X	Trans.	ρ			β			$\hat{\sigma}_{Y X}$
			$\hat{\rho}$	67% CI	90% CI	$\hat{\beta}$	67% CI	90% CI	
CPI Infl.	PCE Infl.	L,L	0.98	0.96, 0.99	0.95, 0.99	1.13	1.07, 1.19	0.98, 1.24	0.36
3M rates	PCE Infl.	L,L	0.47	0.21, 0.80	-0.00, 0.91	0.74	0.34, 1.51	-0.02, 1.89	2.24
10Y rates	PCE Infl.	L,L	0.48	0.21, 0.85	-0.01, 0.92	0.71	0.33, 1.46	0.02, 1.82	2.19
10Y real rates	3M real rates	L,L	0.96	0.90, 0.97	0.75, 0.98	0.97	0.87, 1.11	0.67, 1.31	0.62
CPI Inflation	Money Supply	L,G	0.12	-0.17, 0.55	-0.55, 0.75	0.12	-0.15, 0.50	-0.55, 0.91	2.50
Un. Rate	PCE Infl.	L,L	0.26	-0.03, 0.55	-0.27, 0.80	0.21	-0.04, 0.45	-0.24, 0.83	1.47
Un. Rate	TFP	L,G	-0.65	-0.75, -0.34	-0.91, -0.13	-0.99	-1.38, -0.61	-1.65, -0.27	1.07
3M real rates	Consumption	L,G	0.42	0.08, 0.60	-0.06, 0.75	0.92	0.34, 1.46	-0.35, 2.59	1.84
Stock returns	Consumption	L,G	0.40	0.07, 0.60	-0.08, 0.80	3.00	0.99, 4.99	-0.47, 8.40	6.21
Stock prices	Dividends	G,G	0.20	-0.05, 0.43	-0.30, 0.75	0.45	-0.16, 1.10	-0.61, 1.76	7.28
Stock prices	Earnings	G,G	0.21	-0.04, 0.42	-0.27, 0.57	0.38	-0.14, 0.93	-0.51, 1.35	7.18
Exchange rates	Rel. price ind.	G,G	0.14	-0.13, 0.50	-0.41, 0.75	0.24	-0.37, 0.81	-0.99, 1.26	3.32

Notes: The column labeled "Trans." indicates the transformation applied to the data with "L" denoting level (no transformation) and "G" denoting growth rate (in percentage points at annual rate using scaled first-differences of logarithms). Thus, the levels of inflation, interest rates, and the unemployment were used, and other variables were transformed to growth rates. The long-run regressions were computed from the regression of the variable labeled Y onto the variable labeled X .

figure and table show the long-run covariation of inflation and short- and long-term nominal interest rates. (As above, long-term rates are for 10-year U.S. Treasury bonds available only since 1953, and the analysis with this rate uses $q = 11$). The well-known Fisher relation (Fisher (1930)) decomposes nominal rates into an inflation and real interest rate component making it interesting to gauge how much of the long-run variation in nominal rates can be explained by long-run variation in inflation. The long-run correlation of nominal interest rates and inflation is estimated to be approximately 0.5, although the confidence intervals indicate substantial uncertainty. A unit long-run regression coefficient of nominal rates onto inflation is consistent with data, but the confidence intervals are wide.¹¹ The next entry in the figure and table shows the long-run covariation in short- and long-term real interest rates (constructed as nominal rates minus the PCE inflation rate). Like their nominal counterparts, short- and long-term real rates are highly correlated over the long-run (90% confidence interval: $0.75 \leq \rho \leq 0.98$) with a near unit regression coefficient of long rates onto short rates.

Money growth and inflation: An important implication of the quantity theory of money is the close relationship between money growth and price inflation over the long-run. Lucas (1980) investigated this implication using time series data on money (M1) growth and (CPI) inflation for the U.S. over 1953-1977. After using an exponential smoothing filter to isolate long-run variation in the series, he found a nearly one-for-one relationship between money growth and inflation. The next entry in the figure and table examines this long-run relation using the same M1 and CPI data used by Lucas, but over the longer sample period, 1947-2015. Figure 5 shows the close long-run relationship between money growth and inflation from the mid-1950s through late 1970s documented by Lucas, but shows a much weaker (or non-existent) relationship in the post-1980 sample period, and over the entire sample period the estimated long-run correlation is only 0.12 with a 67% confidence interval that ranges from -0.17 to 0.55.

Long-run Phillips correlation: The next entry summarizes the long-run correlation between the unemployment and inflation. The estimated long-run Phillips correlation and slope coefficient are positive, but $\rho = \beta = 0$ is contained in the 67% confidence interval.

¹¹These estimates measure the long-run Fisher “correlation,” not the long-run Fisher “effect”. The long-run Fisher correlation considers variation from all sources, while the Fisher effect instead considers variation associated with exogenous long-run nominal shocks (e.g., Fisher and Seater (1993), King and Watson (1997)). A similar distinction holds for the Phillips correlation and the Phillips curve (see King and Watson (1994)).

That said, the confidence intervals are wide so that, like the Fisher correlation, the data are not very informative about the long-run Phillips correlation.

Unemployment and productivity: Panel (g) of the figure investigates the long-run co-variation of the unemployment rate and productivity growth. The large negative in-sample long-run correlation evident in the figure has been noted previously (e.g., Staiger, Stock, and Watson (2001)); the confidence intervals reported in Table 7 show that the correlation is unlikely to be spurious. There is a statistically significant negative long-run relationship between the variables. A long-run one percentage point increase in the rate of growth of productivity is associated with an estimated one percentage point decline in the long-run unemployment rate. We are unaware of an economically compelling theoretical explanation for the large negative correlation.

Real returns and consumption growth: Consumption-based asset pricing models (e.g., Lucas (1978)) draw a connection between consumption growth (as an indicator of the intertemporal marginal rate of substitution) and asset returns. A large literature has followed Hansen and Singleton (1982, 1983) investigating this relationship, with varying degrees of success. Rose (1988) discusses the puzzling long-run implications of the model when consumption growth follows an $I(0)$ process and real returns are $I(1)$ (also see Neely and Rapach (2008)), but moving beyond the $I(0)$ and $I(1)$ models, it is clear from the empirical results reported above that both consumption growth and real interest rates exhibit substantial long-run variability. The next two entries in the figure and table investigate the long-run covariability between consumption growth and real returns; first using real returns on short-term treasury bills and then using real returns on stocks. Both suggest a moderate positive long-run correlation between real returns and consumptions growth rates, although the confidence interval is wide (90% confidence range from just below zero to 0.80).

Stock Prices, Dividends, and Earnings: Present value models of stock prices imply a close relationship between long-run values of prices, dividends, and earnings (e.g., Campbell and Shiller (1987)). An implication of this long-run relation in a cointegration framework is that dividends, earnings, and stock prices share a common $I(1)$ trend, so that their growth rates are perfectly correlated in the long-run and the dividend-price or price-earning ratio is useful for predicting future stock returns. This latter implication has been widely investigated (see Campbell and Yogo (2006) for analysis and references). The next two entries show the long-run correlation of stock prices with dividends and with earnings.¹² While there is

¹²The data are for the S&P, and are updated versions of the data used in Shiller (2000) available

considerable uncertainty about the value of the long-run correlation between stock prices and dividends or earnings, the data suggest that the correlation is not strong. For example, values above $\rho = 0.43$ are ruled out by the 67% confidence set and values above 0.75 are ruled out by the 90% sets.

Long-run PPP: The final entry shows results on the long-run correlation between nominal exchange rates (here the U.S. dollar/British pound exchange rate from 1971-2015) and the ratio of nominal prices (here the ratio of CPI indices for the two countries). With the shortened sample, $q = 8$ captures periods longer than 11 years. Long-run PPP implies that the nominal exchange rate should move proportionally with the price ratio over long time spans, so the long-run growth rates of the nominal exchange rate and price ratios should be perfectly correlated. A large literature has tested this proposition in a unit-root and cointegration framework and obtained mixed conclusions. (See Rogoff (1996) and Taylor and Taylor (2004) for discussion and references). From the final row of Table 7, the growth rate of nominal exchange rates and relative nominal prices are only weakly correlated over periods longer than 11 years, although the confidence interval is very wide.

5.3 Longer-run periods

The empirical results shown above relied on projections capturing periods longer than 11 years. While 11 years is longer than typical business cycles, it does incorporate periods corresponding to what some researchers refer to as the “medium run” (Blanchard (1997), Comin and Gertler (2006)). This subsection investigates the robustness of the empirical results to restricting the data to periods longer than 20 years. For the series available over the entire 68-year post-WWII sample period, this entails only the first $q = 6$ cosine transforms ($2T/6 \approx 23$ years); for the 1971-2015 sample for exchange rates and relative prices this entails using $q = 4$ ($2T/4 \approx 22$ years). Results are summarized in Table 8.

The first four sets of entries in Table 8 involve consumption, investment, labor compensation, TFP and GDP. These results are remarkably similar to the results shown in earlier tables. Covariability over periods longer than 20 years is similar to covariability of periods longer than 11 years, although the reduction in information moving from $q = 12$ to $q = 6$ leads to somewhat wider confidence intervals. The other results summarized in Table 7 show much of the same stability, but there are some notable differences. For example, the point

on Robert Shiller’s webpage.

Table 8: Long-run covariation measures for selected variables: periods longer than 20 years

Y	X	ρ			$\hat{\rho}$	β			$\hat{\sigma}_{y x}$
		67% CI	90% CI	$\hat{\beta}$		67% CI	90% CI		
Cons.	GDP		0.89	0.71, 0.95	0.50, 0.97	0.73	0.56, 0.89	0.36, 1.07	0.32
	Inv.		0.56	0.21, 0.79	-0.04, 0.90	1.08	0.50, 1.68	-0.15, 2.29	1.16
$w \times n$	GDP		0.97	0.92, 0.98	0.83, 0.99	1.26	1.15, 1.38	1.04, 1.50	0.24
	TFP		0.71	0.41, 0.94	0.08, 0.96	1.15	0.72, 1.54	0.32, 2.02	0.60
10Y nom. rates	3M nom.rates		0.96	0.92, 0.98	0.84, 0.99	0.97	0.85, 1.09	0.74, 1.20	0.61
10Y real rates	3M real rates		0.94	0.86, 0.97	0.70, 0.98	1.02	0.87, 1.18	0.62, 1.41	0.61
CPI Infl.	PCE Infl.		0.98	0.96, 0.99	0.93, 0.99	1.11	1.05, 1.17	0.97, 1.23	0.28
3M rates	PCE Infl.		0.65	0.30, 0.85	0.00, 0.96	1.02	0.56, 1.58	0.10, 2.18	2.05
10Y rates	PCE Infl.		0.57	0.23, 0.90	-0.03, 0.95	0.92	0.47, 1.44	-0.05, 2.07	2.11
CPI Inflation	Money Supply		0.27	-0.08, 0.59	-0.37, 0.85	0.30	-0.07, 0.67	-0.43, 1.04	2.39
Un. Rate	PCE Infl.		0.27	-0.08, 0.57	-0.36, 0.85	0.23	-0.07, 0.50	-0.39, 0.88	1.34
Un. Rate	TFP		-0.89	-0.95,-0.72	-0.97,-0.57	-1.84	-2.49,-1.41	-2.85,-0.99	0.54
3M real rates	Consumption		0.30	-0.05, 0.59	-0.45, 0.85	0.98	-0.10, 2.07	-1.54, 3.43	1.79
Stock returns	Consumption		0.41	0.00, 0.70	-0.23, 0.90	4.32	0.85, 7.76	-2.88,11.57	5.75
Stock prices	Dividends		0.42	0.02, 0.71	-0.20, 0.85	1.16	0.30, 1.97	-0.48, 2.75	5.58
Stock prices	Earnings		0.23	-0.08, 0.57	-0.37, 0.73	0.64	-0.20, 1.49	-0.95, 2.26	6.12
Exchange rates	Rel. price ind.		0.70	0.27, 0.93	-0.05, 0.96	0.57	0.30, 0.83	-0.02, 1.14	0.99

Notes: This table summarizes the long-run covariance confidence sets for (X, Y) for periods longer than 20 years ($q = 6$ for all entries except exchange rates, which uses $q = 4$). See notes to Table 7 for units of the variables.

estimates suggest a somewhat larger Fisher correlation over longer periods than over shorter periods, and the same holds for stock prices and dividends. In both cases however, the confidence intervals remain wide. Exchange rates and relative nominal prices also appear more highly correlated using these longer periods. And, the puzzling negative correlation between the unemployment rate and TFP appears to be stronger when the sample is restricted to periods longer than 20 years.¹³

6 Concluding remarks

This paper began by highlighting two fundamental problems for inference about long-run variability and covariability: the first was the paucity of sample inference, and the second was that inference critically depends on the data's long-run persistence. Both play an important role in our analysis. The limited sample information is captured by focusing attention on a small number of low-frequency weighted averages that summarize the data's long-run variability and covariability. In large samples and for a reasonably wide range of persistent processes, these weighted averages are approximately normally distributed. Long-run persistence is important, but as we show, only through its effect on the spectral density in a narrow low-frequency band. Using a flexible parameterization of the spectrum for these low frequencies, small-sample normal inference methods allow us to construct asymptotically efficient confidence intervals for long-run variance and covariance parameters.

This paper has focused on inference about long-run covariability of two time series. Just as with previous frameworks, it is natural to consider a generalization to a higher dimensional setting. For example, this would allow one to determine whether the significant long-run correlation between the unemployment rate and productivity is robust to including a control for, say, some measure of human capital accumulation.

Many elements of our analysis generalize to n time series in a straightforward manner: The analogous definition of Ω_T is equally natural as a second-moment summary of the covariability of n series, and gives rise to corresponding regression parameters, such as coefficients from a $n - 1$ dimensional multiple regression, corresponding residual standard deviations

¹³This analysis has focused on long-run covariability associated with the first q cosine transforms, that is the projections for periods between $2T$ and $2T/q$. More generally, a researcher might be interested in the covariance for periods between, say, $2T/q_1$ and $2T/q_2$ associated with the projections of the data onto Ψ_{q_1} through Ψ_{q_2} . See the appendix for discussion.

and population R^2 s.¹⁴ Multivariate versions of Ω_T can also be used for long-run instrumental variable regressions. As shown in Müller and Watson (forthcoming), the Central Limit Theorem that reduces the inference question to one about the covariance matrix of a multivariate normal holds for arbitrary fixed n . The (A, B, c, d) model of persistence naturally generalizes to a n dimensional system.

Having said that, our numerical approach for constructing (approximate) minimal-length confidence sets faces daunting computational challenges in a higher order system: The quadratic forms that determine the likelihood require $O(n^2q^2)$ floating point operations. Worse still, even for n as small as $n = 3$, the number of parameters in the (A, B, c, d) model is equal to 21. So even after imposing invariance or equivariance, ensuring coverage requires an exhaustive search over a high dimensional nuisance parameter space.

At the same time, it would seem to be relatively straightforward to determine Bayes credible sets also for larger values of n : Under our asymptotic approximation, the (A, B, c, d) parameters enter the likelihood through the covariance matrix of a $nq \times 1$ multivariate normal, so with some care, modern posterior samplers should be able to reliably determine the posterior for any function of interest. Of course, such an approach does not guarantee frequentist coverage, and the empirical results will depend on the choice of prior in a non-trivial way. In this regard, our empirical results in the bivariate system show an interesting pattern: Especially at a lower nominal coverage level, for many realizations, there is no need to augment the Bayes credible set computed from the bivariate fractional model. This suggests that the frequentist coverage of the unaltered Bayes intervals may be close to the nominal level, so these Bayes sets would not be too misleading even from a frequentist perspective.¹⁵ While this will be difficult to exhaustively check, this pattern might well generalize also to larger values of n .

¹⁴Müller and Watson (forthcoming) provide the details of inference in the $I(0)$ model.

¹⁵In fact, a calculation analogous to those in Table 4 shows that the 67% Bayes set contains the true value of $\rho = 0$ at least 64% of the time in the bivariate (A, B, c, d) model, and the 95% Bayes set has coverage of 83%.

References

- BAXTER, M., AND R. G. KING (1999): “Measuring business cycles: approximate band-pass filters for economic time series,” *Review of economics and statistics*, 81(4), 575–593.
- BLANCHARD, O. J. (1997): “The Medium Run,” *Brooking Panel on Economic Activity*, pp. 89–158.
- BOSKIN, M. J., E. DULBERGER, R. GORDON, Z. GRILICHES, AND D. JORGENSON (1996): “Toward a More Accurate Measure of the Cost of Living,” *Final Report to the Senate Finance Committee*.
- BRILLINGER, D. R. (2001): *Time series: data analysis and theory*. SIAM.
- BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*. Springer, New York, second edn.
- CAMPBELL, J. Y., AND R. J. SHILLER (1987): “Cointegration and Tests of Present Value Models,” *Journal of Political Economy*, 95, 1062–1088.
- CAMPBELL, J. Y., AND M. YOGO (2006): “Efficient Tests of Stock Return Predictability,” *Journal of Financial Economics*, 81, 27–60.
- CHEN, W. W., AND C. M. HURVICH (2003): “Semiparametric estimation of multivariate fractional cointegration,” *Journal of the American Statistical Association*, 98(463), 629–642.
- COMIN, D., AND M. GERTLER (2006): “Medium Term Business Cycles,” *American Economic Review*, 96, 523–551.
- DEN HAAN, W. J., AND A. T. LEVIN (1997): “A Practitioner’s Guide to Robust Covariance Matrix Estimation,” in *Handbook of Statistics 15*, ed. by G. S. Maddala, and C. R. Rao, pp. 299–342. Elsevier, Amsterdam.
- ELLIOTT, G. (1998): “The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots,” *Econometrica*, 66, 149–158.
- ELLIOTT, G., U. K. MÜLLER, AND M. W. WATSON (2015): “Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis,” *Econometrica*, 83, 771–811.

- ENGLE, R. F. (1974): “Band spectrum regression,” *International Economic Review*, 15, 1–11.
- FISHER, I. (1930): *The Theory of Interest*. MacMillan, New York.
- FISHER, M. E., AND J. D. SEATER (1993): “Long-run Neutrality and Superneutrality in an ARIMA Framework,” *American Economic Review*, 83, 402–415.
- GORDON, R. J. (2006): “The Boskin Commission Report: A Retrospective One Decade Later,” *NBER working paper 12311*.
- GRANGER, C. W. J., AND P. NEWBOLD (1974): “Spurious Regressions in Econometrics,” *Journal of Econometrics*, 2, 111–120.
- HANSEN, L. P., AND K. J. SINGLETON (1982): “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica*, 50(5), 1269–1286.
- (1983): “Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns,” *Journal of Political Economy*, 91(2), 249–265.
- HODRICK, R. J., AND E. C. PRESCOTT (1997): “Postwar US business cycles: an empirical investigation,” *Journal of Money, Credit and Banking*, 29, 1–16.
- JANSSON, M., AND M. J. MOREIRA (2006): “Optimal Inference in Regression Models with Nearly Integrated Regressors,” *Econometrica*, 74, 681–714.
- KING, R. G., AND M. W. WATSON (1994): “The Post-War U.S. Phillips Curve: A Revisionist Econometric History,” *Carnegie-Rochester Conference on Public Policy*, 41, 157–219.
- (1997): “Testing long-run neutrality,” *Federal Reserve Bank of Richmond Economic Quarterly*, 83, 69–101.
- KLEIN, L. R., AND R. F. KOSOBUD (1961): “Some Econometrics of Growth: Great Ratios of Economics,” *Quarterly Journal of Economics*, LXXV, 173–198.
- LINDGREN, G. (2013): *Stationary Stochastic Processes: Theory and Applications*. Chapman & Hall/CRC.

- LUCAS, R. E. (1978): “Asset Prices in an Exchange Economy,” *Econometrica*, 46(6), 1429–1445.
- (1980): “Two Illustrations of the Quantity Theory of Money,” *American Economic Review*, 70(5), 1005–1014.
- MARINUCCI, D., AND P. M. ROBINSON (2001): “Semiparametric fractional cointegration analysis,” *Journal of Econometrics*, 105(1), 225–247.
- MÜLLER, U. K. (2011): “Efficient Tests under a Weak Convergence Assumption,” *Econometrica*, 79, 395–435.
- MÜLLER, U. K., AND A. NORETS (2016): “Credibility of Confidence Sets in Nonstandard Econometric Problems,” *Econometrica*, 84, 2183–2213.
- MÜLLER, U. K., AND M. W. WATSON (2008): “Testing Models of Low-Frequency Variability,” *Econometrica*, 76, 979–1016.
- (2016): “Measuring Uncertainty about Long-Run Predictions,” *Review of Economic Studies*, 83.
- (forthcoming): “Low-Frequency Econometrics,” in *Advances in Economics: Eleventh World Congress of the Econometric Society*, ed. by B. Honoré, and L. Samuelson. Cambridge University Press.
- NEELY, C. J., AND D. E. RAPACH (2008): “Real Interest Rate Persistence: Evidence and Implications,” *Federal Reserve Bank of St. Louis Review*, 90(6), 609–641.
- PHILLIPS, P. C. B. (1991): “Spectral Regression for Cointegrated Time Series,” in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, and G. Tauchen, p. 4. Cambridge University Press, Cambridge.
- PRATT, J. W. (1961): “Length of Confidence Intervals,” *Journal of the American Statistical Association*, 56, 549–567.
- ROBINSON, P. M. (2008): “Multiple local Whittle estimation in stationary systems,” *The Annals of Statistics*, 36(5), 2508–2530.

- ROBINSON, P. M., AND J. HUALDE (2003): “Cointegration in fractional systems with unknown integration orders,” *Econometrica*, 71(6), 1727–1766.
- ROGOFF, K. (1996): “The Purchasing Power Parity Puzzle,” *Journal of Economic Literature*, 34, 647–668.
- ROSE, A. K. (1988): “Is the Real Interest Rate Stable?,” *Journal of Finance*, 43(5), 1095–1112.
- SHILLER, R. J. (2000): *Irrational Exuberance*. Princeton University Press, Princeton.
- SHIMOTSU, K. (2012): “Exact local Whittle estimation of fractionally cointegrated systems,” *Journal of Econometrics*, 169(2), 266–278.
- STAIGER, D., J. H. STOCK, AND M. W. WATSON (2001): “Prices, Wages, and the U.S. NAIRU in the 1990s,” in *The Roaring Nineties*, ed. by A. B. Krueger, and R. Solow, pp. 3–60, New York. The Russell Sage Foundation.
- TAYLOR, A. M., AND M. P. TAYLOR (2004): “The Purchasing Power Parity Debate,” *Journal of Economic Perspectives*, 18, 135–158.

Supplementary Appendix to

Long-Run Covariability

(This version: December 2017)

by Ulrich K. Müller and Mark W. Watson

This appendix provides supplemental material. Section A.1 discusses the form of the confidence sets; Section A.2 derives the necessary densities; Section A.3 discusses the numerically determined approximate least favorable distributions; the data are described in Section A.4; Section A.5 compares low-pass and low-frequency projections for GDP and consumption; Section A.6 discusses alternative versions of Ω_T constructed from projections onto a subset of the columns of Ψ_T and summarizes the resulting empirical results.

A.1 Form of Confidence Sets

For each of the three sets H^ρ , H^β and H^σ , we exogenously impose that they contain the $(1 - \alpha)$ equal-tailed invariant credible set relative to the prior F , as suggested by Müller and Norets (2016). Denote this credible set by H_0^i , $i \in \{\rho, \beta, \sigma\}$. Specializing Theorem 3 of Müller and Norets (2016) to the three problems considered here yields the following form for the three type of confidence sets:

H^ρ : Let $X^s = X/\sqrt{X'X}$ and $Y^s = Y/\sqrt{Y'Y}$, and let $f^s(x^s, y^s|\theta)$ be the density of (X^s, Y^s) under $\theta \in \Theta$.¹⁶ Then

$$\begin{aligned} H_0^\rho(x, y) &= \left\{ r : \alpha/2 \leq \frac{\int \mathbf{1}[g^\rho(\theta) \leq r] f^s(x^s, y^s|\theta) dF(\theta)}{\int f^s(x^s, y^s|\theta) dF(\theta)} \leq 1 - \alpha/2 \right\} \\ H^\rho(x, y) &= \left\{ r : \int f^s(x^s, y^s|\theta) dW(\theta) \geq \int f^s(x^s, y^s|\theta) d\Lambda_r^\rho(\theta) \right\} \cup H_0^\rho(x, y) \end{aligned} \quad (\text{A.1})$$

where W is the weighting function over which expected length is minimized and the family of positive measures Λ_r^ρ on Θ , indexed by $r \in (-1, 1)$, are such that $\Lambda_r^\rho(\{\theta : g^\rho(\theta) \neq r \text{ or } P_\theta(g^\rho(\theta) \in H^\rho(X, Y)) > 1 - \alpha\}) = 0$ and $P_\theta(g^\rho(\theta) \in H^\rho(X, Y)) \geq 1 - \alpha$ for all $\theta \in \Theta$.

¹⁶Here and in the following, we distinguish between random variables and generic real numbers by the usual upper case / lower case convention. We also implicitly assume the same functional relationship between the random variables and their corresponding real variables, if appropriate. For example, (x^s, y^s) on the right hand side of (A.1) is implicitly thought of as a function of (x, y) .

H^β : Let the $q-2$ vectors X^* and Y^* , and $X_0^*, U_{11}, U_{12}, U_{22} \in \mathbb{R}$ be such that

$$(X, Y) = \left(\begin{pmatrix} 1 \\ X_0^* \\ X^* \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ Y^* \end{pmatrix} \right) \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix},$$

that is, perform the LDU decomposition of the upper 2×2 block of the $q \times 2$ matrix (X, Y) .

Let $Z^* = (X_0^*, X^*, Y^*)'$. Then

$$\begin{aligned} H_0^\beta(x, y) &= \left\{ b : \alpha/2 \leq \frac{\int \int \mathbf{1}[\frac{u_{11}b - u_{12}}{u_{22}} \leq w] f_0^\beta(z^*, w|\theta) dw dF(\theta)}{\int f_1^*(z^*|\theta) dF(\theta)} \leq 1 - \alpha/2 \right\} \\ H^\beta(x, y) &= \left\{ b : \int h^\beta(z^*|\theta) f_1^*(z^*|\theta) dW(\theta) \geq \int f_0^\beta(z^*, \frac{u_{11}b - u_{12}}{u_{22}}|\theta) d\Lambda^\beta(\theta) \right\} \cup H_0^\beta(x, y) \end{aligned}$$

where $f_1^*(z^*|\theta)$ is the density of Z^* under θ , $h^\beta(z^*|\theta) = E_\theta[|U_{22}/U_{11}| | Z^* = z^*]$, $f_0^\beta(z^*, w|\theta)$ is the density of the $2q-2$ vector $(Z^{*'}, (U_{11}g^\beta(\theta) - U_{12})/U_{22})'$ under θ , and Λ^β is a positive measure on Θ such that $\Lambda^\beta(\{\theta : P_\theta(g^\beta(\theta) \in H^\beta(X, Y)) > 1 - \alpha\}) = 0$ and $P_\theta(g^\beta(\theta) \in H^\beta(X, Y)) \geq 1 - \alpha$ for all $\theta \in \Theta$.

H^σ :

$$\begin{aligned} H_0^\sigma(x, y) &= \left\{ s : \alpha/2 \leq \frac{\int \int \mathbf{1}[\frac{s}{u_{22}} \leq w] f_0^\sigma(z^*, w|\theta) dw dF(\theta)}{\int f_1^*(z^*|\theta) dF(\theta)} \leq 1 - \alpha/2 \right\} \\ H^\sigma(x, y) &= \left\{ s : \int h^\sigma(z^*|\theta) f_1^*(z^*|\theta) dW(\theta) \geq \int f_0^\sigma(z^*, \frac{s}{u_{22}}|\theta) d\Lambda^\sigma(\theta) \right\} \cup H_0^\sigma(x, y) \end{aligned}$$

where $h^\sigma(z^*|\theta) = E[|U_{22}| | Z^* = z^*]$ under θ , $f_0^\sigma(z^*, w|\theta)$ is the density of the $2q-2$ vector $(Z^{*'}, g^\sigma(\theta)/|U_{22}|)'$ under θ , and Λ^σ is a positive measure on Θ such that $\Lambda^\sigma(\{\theta : P_\theta(g^\sigma(\theta) \in H^\sigma(X, Y)) > 1 - \alpha\}) = 0$ and $P_\theta(g^\sigma(\theta) \in H^\sigma(X, Y)) \geq 1 - \alpha$ for all $\theta \in \Theta$.

It remains to derive f^s , f_1^* , $f_1^* h^\beta$, $f_1^* h^\sigma$, f_0^β and f_0^σ , and to determine Λ_r^ρ , Λ^β and Λ^σ .

A.2 Densities of Maximal Invariants and Related Results

A.2.1 Preliminaries

As we show below, most densities of interest involve integrals of the form

$$Q(r) = \frac{1}{2\pi} \int_0^\infty \int_0^\infty s^{p_1} t^{p_2} \exp[-\frac{1}{2} \begin{pmatrix} s \\ t \end{pmatrix}' \begin{pmatrix} a^2 & abr \\ abr & b^2 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}] ds dt$$

$$\begin{aligned}
&= a^{p_1-1}b^{p_2-1}\frac{1}{2\pi}\int_0^\infty\int_0^\infty s^{p_1}t^{p_2}\exp[-\frac{1}{2}(s^2-2rst+t^2)]dsdt \\
&= a^{p_1-1}b^{p_2-1}\int_0^\infty t^{p_2}\phi(\sqrt{1-r^2}t)\int_0^\infty s^{p_1}\phi(s-rt)dsdt
\end{aligned}$$

for nonnegative integers p_1 and p_2 , positive reals a, b , and $-1 < r < 1$, with ϕ the p.d.f. of a standard normal distribution. Note that

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^\infty\int_{-\infty}^\infty |s|^{p_1}|t|^{p_2}\exp[-\frac{1}{2}\begin{pmatrix} s \\ t \end{pmatrix}'\begin{pmatrix} a^2 & abr \\ abr & b^2 \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix}]dsdt \\
&= 2Q(r) + 2Q(-r).
\end{aligned}$$

We initially discuss how to obtain closed-form expressions for $Q(r)$. The resulting explicit formulae for densities, even after simplification with a computer algebra system, are long and uninformative, and they are relegated to the replication files.

Lemma A.1 *Let Φ be the c.d.f. of a standard normal random variable.*

(a) *With $B_p(m) = \int_{-\infty}^\infty \phi(s-m)s^p ds$, we have $B_0(m) = 1$, $B_1(m) = m$ and*

$$B_{p+2}(m) = (p+1)B_p(m) + mB_{p+1}(m);$$

(b) *With $I_l(h) = \frac{1}{\Phi(h)}\int_{-\infty}^h \phi(s)s^l ds$,*

$$\int_{-\infty}^0 \phi(s+h)s^p ds = \Phi(h)\sum_{l=0}^p \binom{p}{l}(-h)^{p-l}I_l(h)$$

and $I_0(h) = 1$, $I_1(h) = -\phi(h)/\Phi(h)$ and $I_p(h) = -h^{p-1}\phi(h)/\Phi(h) + (p-1)I_{p-2}(h)$;

(c) $\sqrt{2\pi}\int_0^\infty \phi(\sqrt{1+c^2}s)s^{p+1}ds = 2^{\frac{p}{2}}\Gamma(1+p/2)(1+c^2)^{-p/2-1}$;

(d) *With $A_p(r) = 2\pi\int_0^\infty \phi(s)\Phi(\frac{r}{\sqrt{1-r^2}}s)s^p ds$, $A_0(r) = \pi - \arccos(r)$, $A_1(r) = \sqrt{\pi/2}(1+r)$,*

and

$$A_{p+2}(r) = (p+1)A_p(r) + \Gamma(1+p/2)2^{p/2}r(1-r^2)^{(1+p)/2}.$$

Proof. (a) By integration by parts and $\phi'(s) = -s\phi(s)$

$$\int_{-\infty}^\infty \phi(s-m)s^p ds = \int_{-\infty}^\infty (s-m)\phi(s-m)\frac{s^{p+1}}{p+1}ds$$

and the result follows.

(b) See Dhrymes (2005).

(c) Immediate after substituting $s^2 \rightarrow u$ from the definition of the Gamma function.

(d) Define $\tilde{A}_p(c) = 2\pi \int_0^\infty \phi(s)\Phi(cs)s^p ds$, so that $A_p(r) = \tilde{A}_p(r/\sqrt{1-r^2})$. Note that $\tilde{A}_p(0) = \pi \int_0^\infty \phi(s)s^p ds$, and $\tilde{A}'_p(c) = d\tilde{A}_p(c)/dc = 2\pi \int_0^\infty \phi(s)\phi(cs)s^{p+1} ds = \sqrt{2\pi} \int_0^\infty \phi(\sqrt{1+c^2}s)s^{p+1} ds$. Now $\tilde{A}_p(c) = \tilde{A}_p(0) + \int_0^c \tilde{A}'_p(u)du$. The results for $A_0(r)$ and $A_1(r)$ now follow by applying (c) and a direct calculation. For the iterative expression, by integration by parts and $\phi'(s) = -s\phi(s)$

$$\begin{aligned}\tilde{A}_p(c) &= \left[2\pi\phi(s)\Phi(cs)\frac{s^{p+1}}{p+1} \right]_0^\infty - 2\pi \int_0^\infty \frac{s^{p+1}}{p+1} (c\phi(cs)\phi(s) - s\phi(s)\Phi(cs)) ds \\ &= \frac{1}{p+1} \left(\tilde{A}_{p+2}(c) - c\sqrt{2\pi} \int_0^\infty \phi(\sqrt{1+c^2}s)s^{p+1} ds \right),\end{aligned}$$

and the result follows from applying part (c). ■

Now by Lemma A.1 (a),

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty s^{p_1} \exp[-\frac{1}{2}(s-rt)^2] ds = C_0(rt) - \int_{-\infty}^0 \phi(s-rt)s^{p_1} ds$$

for some polynomial C_0 whose coefficients may be determined explicitly by the formula in Lemma A.1 (a). Furthermore,

$$\int_{-\infty}^0 \phi(s-rt)s^{p_1} ds = \phi(rt)C_1(rt) + \Phi(rt)C_2(t)$$

for some polynomials C_1 and C_2 that may be determined explicitly by the formula in Lemma A.1 (b). The remaining integral over dt is of the form

$$\begin{aligned}& \int_0^\infty t^{p_2} \phi(\sqrt{1-r^2}t) [C_0(rt) - \phi(rt)C_1(rt) - \Phi(rt)C_2(t)] dt \\ &= (1-r^2)^{p_2/2-1} \int_0^\infty \phi(t)t^{p_2} C_0\left(\frac{r}{\sqrt{1-r^2}}t\right) dt - \frac{1}{\sqrt{2\pi}} \int_0^\infty \phi(t)t^{p_2} C_1(rt) dt \\ & \quad - (1-r^2)^{p_2/2-1} \int_0^\infty \phi(t)\Phi\left(\frac{r}{\sqrt{1-r^2}}t\right)t^{p_2} C_2\left(\frac{r}{\sqrt{1-r^2}}t\right) dt\end{aligned}$$

which can be determined explicitly by applying Lemma A.1 (c)-(d).

In the following, we simply write Σ for the covariance matrix of $\text{vec}(X, Y)$, keeping the dependence on θ implicit. If not specified otherwise, all integrals are over the entire real line. Also, denote the four $q \times q$ blocks of the inverse of Σ as

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{xx}^- & \Sigma_{xy}^- \\ \Sigma_{yx}^- & \Sigma_{yy}^- \end{pmatrix}.$$

A.2.2 Derivation of f^s

Let $S_x = \sqrt{X'X}$ and $S_y = \sqrt{Y'Y}$. Write μ_l for Lebesgue measure on \mathbb{R}^l , and ν_q for the surface measure of a q dimensional unit sphere. For $x \in \mathbb{R}^q$, let $x = x^s s_x$, where x^s is a point on the surface of a q dimensional unit sphere, and $s_x \in \mathbb{R}^+$. By Theorem 2.1.13 of Muirhead (1982), $d\mu_q(x) = s_x^{q-1} d\nu_q(x^s) d\mu_1(s_x)$. We thus can write the joint density of (X^s, Y^s, S_x, S_y) with respect to $\nu_q \times \nu_q \times \mu_1 \times \mu_1$ as

$$\begin{aligned} & (2\pi)^{-q} (\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2} \begin{pmatrix} x^s s_x \\ y^s s_y \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} x^s s_x \\ y^s s_y \end{pmatrix}\right] s_x^{q-1} s_y^{q-1} \\ &= (2\pi)^{-q} (\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2} \begin{pmatrix} s_x \\ s_y \end{pmatrix}' \begin{pmatrix} x^{s'} \Sigma_{xx}^- x^s & x^{s'} \Sigma_{xy}^- y^s \\ y^{s'} \Sigma_{yx}^- x^s & y^{s'} \Sigma_{yy}^- y^s \end{pmatrix} \begin{pmatrix} s_x \\ s_y \end{pmatrix}\right] s_x^{q-1} s_y^{q-1} \end{aligned}$$

and the marginal density of $(X^{s'}, Y^{s'})'$ with respect to $\nu_q \times \nu_q$ is

$$(2\pi)^{-q} (\det \Sigma)^{-1/2} \int_0^\infty \int_0^\infty \exp\left[-\frac{1}{2} \begin{pmatrix} s_x \\ s_y \end{pmatrix}' \begin{pmatrix} x^{s'} \Sigma_{xx}^- x^s & x^{s'} \Sigma_{xy}^- y^s \\ y^{s'} \Sigma_{yx}^- x^s & y^{s'} \Sigma_{yy}^- y^s \end{pmatrix} \begin{pmatrix} s_x \\ s_y \end{pmatrix}\right] s_x^{q-1} s_y^{q-1} ds_x ds_y.$$

A.2.3 Derivation of f_1^*

With $X^\dagger = (1, X_0^*, X^*)'$, $Y^\dagger = (1, 0, Y^{*'})'$ and $U = \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix}$, we have

$$\begin{aligned} (X, Y) &= (X^\dagger, Y^\dagger) U \\ &= \begin{pmatrix} U_{11} & U_{12} \\ U_{11} X_1^* & U_{12} X_1^* + U_{22} \\ U_{11} X^* & U_{12} X^* + U_{22} Y^* \end{pmatrix}. \end{aligned}$$

This equation, viewed as a $\mathbb{R}^{2q} \rightarrow \mathbb{R}^{2q}$ function of $T^* = (X^*, Y^*, X_0^*, U_{11}, U_{12}, U_{22})$ has Jacobian determinant $U_{11}^{q-1} U_{22}^{q-2}$, so that the density of T^* is

$$f_{T^*}(t^*) = (2\pi)^{-q} (\det \Sigma)^{-1/2} |u_{11}|^{q-1} |u_{22}|^{q-2} \exp\left[-\frac{1}{2} (\text{vec } z^\dagger u)' \Sigma^{-1} (\text{vec } z^\dagger u)\right] \quad (\text{A.2})$$

with $z^\dagger = (x^\dagger, y^\dagger)$, and we are left to integrate out u_{11} , u_{12} and u_{22} . Using $\text{vec}(z^\dagger u) = (I_2 \otimes z^\dagger) \text{vec}(u)$, we have

$$\text{vec}(z^\dagger u)' \Sigma^{-1} \text{vec}(z^\dagger u) = \text{vec}(u)' [(I_2 \otimes z^\dagger)' \Sigma^{-1} (I_2 \otimes z^\dagger)] \text{vec}(u)$$

$$\begin{aligned}
&= \begin{pmatrix} u_{11} \\ 0 \\ u_{12} \\ u_{22} \end{pmatrix}' \begin{pmatrix} z^\dagger & 0 \\ 0 & z^\dagger \end{pmatrix}' \begin{pmatrix} \Sigma_{xx}^- & \Sigma_{xy}^- \\ \Sigma_{yx}^- & \Sigma_{yy}^- \end{pmatrix} \begin{pmatrix} z^\dagger & 0 \\ 0 & z^\dagger \end{pmatrix} \begin{pmatrix} u_{11} \\ 0 \\ u_{12} \\ u_{22} \end{pmatrix} \\
&= \begin{pmatrix} u_{11} \\ u_{22} \\ u_{12} \end{pmatrix}' \begin{pmatrix} x^{\dagger'} \Sigma_{xx}^- x^\dagger & \cdot & \cdot \\ y^{\dagger'} \Sigma_{yx}^- x^\dagger & y^{\dagger'} \Sigma_{yy}^- y^\dagger & \cdot \\ x^{\dagger'} \Sigma_{yx}^- x^\dagger & y^{\dagger'} \Sigma_{yy}^- y^\dagger & x^{\dagger'} \Sigma_{yy}^- x^\dagger \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{22} \\ u_{12} \end{pmatrix} \\
&= \begin{pmatrix} \hat{u} \\ u_{12} \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} \hat{u} \\ u_{12} \end{pmatrix}
\end{aligned}$$

where $\hat{u} = (u_{11}, u_{22})'$, $V = \begin{pmatrix} x^{\dagger'} \Sigma_{xx}^- x^\dagger & y^{\dagger'} \Sigma_{yx}^- x^\dagger \\ y^{\dagger'} \Sigma_{yx}^- x^\dagger & y^{\dagger'} \Sigma_{yy}^- y^\dagger \end{pmatrix}$, $v' = (x^{\dagger'} \Sigma_{yx}^- x^\dagger, y^{\dagger'} \Sigma_{yy}^- x^\dagger)$ and $v_0^2 = x^{\dagger'} \Sigma_{yy}^- x^\dagger$. Furthermore, by “completing the square”,

$$\int \exp[-\frac{1}{2} \begin{pmatrix} \hat{u} \\ u_{12} \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} \hat{u} \\ u_{12} \end{pmatrix}] du_{12} = \sqrt{2\pi} v_0^{-1} \exp[-\frac{1}{2} \hat{u}' (V - v v' / v_0) \hat{u}]$$

and with $\tilde{V} = V - v v' / v_0$, we obtain

$$f_1^*(z^*) = (2\pi)^{-q+1/2} (\det \Sigma)^{-1/2} v_0^{-1} \int \int |u_{11}|^{q-1} |u_{22}|^{q-2} \exp[-\frac{1}{2} \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}' \tilde{V} \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}] du_{11} du_{22}.$$

A.2.4 Derivation of $f_1^* h^\beta$

We have

$$\begin{aligned}
h^\beta(z^*|\theta) &= E_\theta[|U_{22}/U_{11}| | Z^* = z^*] \\
&= \frac{\int \int \int |\frac{u_{22}}{u_{11}}| f_{T^*}(t^*) du_{12} du_{11} du_{22}}{f_1^*(z^*)}.
\end{aligned}$$

Thus, proceeding as in the derivation of f_1^* yields

$$\begin{aligned}
&h^\beta(z^*|\theta) f_1^*(z^*|\theta) \\
&= \int \int \int |\frac{u_{22}}{u_{11}}| f_{T^*}(t^*) du_{12} du_{11} du_{22} \\
&= (2\pi)^{-q+1/2} (\det \Sigma)^{-1/2} v_0^{-1} \int \int |u_{11}|^{q-2} |u_{22}|^{q-1} \exp[-\frac{1}{2} \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}' \tilde{V} \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}] du_{11} du_{22}.
\end{aligned}$$

A.2.5 Derivation of $f_1^* h^\sigma$

Proceeding analogously to the derivation of $f_1^* h^\beta$, we obtain

$$\begin{aligned}
& h^\sigma(z^*|\theta) f_1^*(z^*|\theta) \\
&= \int \int \int |u_{22}| f_{T^*}(t^*) du_{12} du_{11} du_{22} \\
&= (2\pi)^{-q+1/2} (\det \Sigma)^{-1/2} v_0^{-1} \int \int |u_{11}|^{q-1} |u_{22}|^{q-1} \exp[-\frac{1}{2} \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}' \tilde{V} \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}] du_{11} du_{22}.
\end{aligned}$$

A.2.6 Derivation of f_0^β

With $W^\beta = (U_{11}g^\beta(\theta) - U_{12})/U_{22}$, we have

$$\begin{pmatrix} U_{11} \\ U_{22} \\ U_{12} \end{pmatrix} = \begin{pmatrix} U_{11} \\ U_{22} \\ U_{11}g^\beta(\theta) - U_{22}W^\beta \end{pmatrix} = \begin{pmatrix} \hat{U} \\ \lambda'_W \hat{U} \end{pmatrix}$$

with $\hat{U} = (U_{11}, U_{22})'$ and $\lambda_W = (g^\beta(\theta), -W^\beta)'$. This equation, viewed as $\mathbb{R}^3 \mapsto \mathbb{R}^3$ function of $(U_{11}, U_{22}, W^\beta)$, has Jacobian determinant equal to $-U_{22}$. Thus, with $u_w = \begin{pmatrix} u_{11} & \lambda'_w \hat{u} \\ 0 & u_{22} \end{pmatrix}$, the joint density of (Z^*, W^β) can be written as

$$\int \int (2\pi)^{-q} (\det \Sigma)^{-1/2} |u_{11}|^{q-1} |u_{22}|^{q-1} \exp[-\frac{1}{2} (\text{vec } z^\dagger u_w)' \Sigma^{-1} (\text{vec } z^\dagger u_w)] du_{11} du_{22}.$$

Now similar to the derivation of f_1^* ,

$$\begin{aligned}
(\text{vec } z^\dagger u_w)' \Sigma^{-1} (\text{vec } z^\dagger u_w) &= \begin{pmatrix} \hat{u} \\ \lambda'_w \hat{u} \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} \hat{u} \\ \lambda'_w \hat{u} \end{pmatrix} \\
&= \hat{u}' \begin{pmatrix} I_2 \\ \lambda'_w \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} I_2 \\ \lambda'_w \end{pmatrix} \hat{u}.
\end{aligned}$$

Thus, with $V_w = \begin{pmatrix} I_2 \\ \lambda'_w \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} I_2 \\ \lambda'_w \end{pmatrix}$,

$$f_0^\beta(z^*, w|\theta) = (2\pi)^{-q} (\det \Sigma)^{-1/2} \int \int |u_{11}|^{q-1} |u_{22}|^{q-1} \exp[-\frac{1}{2} \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}' V_w \begin{pmatrix} u_{11} \\ u_{22} \end{pmatrix}] du_{11} du_{22}.$$

A.2.7 Derivation of f_0^σ

Let $\tilde{W}^\sigma = g^\sigma(\theta)/U_{22}$, so that $W^\sigma = |\tilde{W}^\sigma|$. Let $\tilde{f}_0^\sigma(z^*, \tilde{w}^\sigma|\theta)$ be the joint density of (Z^*, \tilde{W}^σ) . Then $f_0^\sigma(z^*, w^\sigma|\theta) = \tilde{f}_0^\sigma(z^*, w^\sigma|\theta) + \tilde{f}_0^\sigma(z^*, -w^\sigma|\theta)$, so it suffices to derive an expression for \tilde{f}_0^σ .

We have

$$\begin{pmatrix} U_{11} \\ U_{22} \\ U_{12} \end{pmatrix} = \begin{pmatrix} U_{11} \\ g^\sigma(\theta)/\tilde{W}^\sigma \\ U_{12} \end{pmatrix}.$$

This equation, viewed as $\mathbb{R}^3 \mapsto \mathbb{R}^3$ function of $(U_{11}, U_{12}, \tilde{W}^\sigma)$, has Jacobian determinant equal to $-g^\sigma(\theta)/\tilde{W}^{\sigma 2}$. From (A.2), with $u_w^\sigma = \begin{pmatrix} u_{11} & u_{12} \\ 0 & g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}$, the joint density of (Z^*, \tilde{W}^σ) can thus be written as

$$(2\pi)^{-q}(\det \Sigma)^{-1/2}|\tilde{w}^\sigma|^{-q}|g^\sigma(\theta)|^{q-1} \int \int |u_{11}|^{q-1} \exp[-\frac{1}{2}(\text{vec } z^\dagger u_w^\sigma)' \Sigma^{-1} (\text{vec } z^\dagger u_w^\sigma)] du_{12} du_{11}.$$

Now similar to the derivation of f_1^* ,

$$(\text{vec } z^\dagger u_w^\sigma)' \Sigma^{-1} (\text{vec } z^\dagger u_w^\sigma) = \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \\ u_{12} \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \\ u_{12} \end{pmatrix}$$

and

$$\begin{aligned} & \int \exp[-\frac{1}{2} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \\ u_{12} \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \\ u_{12} \end{pmatrix}] du_{12} \\ &= \sqrt{2\pi} v_0^{-1} \exp[-\frac{1}{2} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}' (V - vv'/v_0) \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}] \end{aligned}$$

so that

$$\begin{aligned} \tilde{f}_0^\sigma(z^*, \tilde{w}^\sigma|\theta) &= (2\pi)^{-q+1/2}(\det \Sigma)^{-1/2}|g^\sigma(\theta)|^{q-1}|\tilde{w}^\sigma|^{-q}v_0^{-1} \\ &\quad \times \int |u_{11}|^{q-1} \exp[-\frac{1}{2} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}' \tilde{V} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}] du_{11}. \end{aligned}$$

Furthermore, with \tilde{v}_{ij}^2 the i, j th element of \tilde{V} , $b_w = \tilde{v}_{12}g^\beta(\theta)/(\tilde{w}^\sigma\tilde{v}_{11})$ and $\tilde{a}_w = \tilde{v}_{22}(g^\beta(\theta)/\tilde{w}^\sigma)^2$

$$\begin{aligned}
& \int |u_{11}|^{q-1} \exp[-\tfrac{1}{2} \left(\begin{array}{c} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{array} \right)' \tilde{V} \left(\begin{array}{c} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{array} \right)] du_{11} \\
&= \int |u_{11}|^{q-1} \exp[-\tfrac{1}{2}(\tilde{v}_{11}^2 u_{11}^2 + 2b_w \tilde{v}_{11} u_{11} + a_w^2)] du_{11} \\
&= \tilde{v}_{11}^{-q} \int |\omega|^{q-1} \exp[-\tfrac{1}{2}(\omega^2 + 2b_w \omega + a_w^2)] d\omega \\
&= \tilde{v}_{11}^{-q} \exp[-\tfrac{1}{2}(a_w^2 - b_w^2)] \int |\omega|^{q-1} \exp[-\tfrac{1}{2}(\omega + b_w)^2] d\omega.
\end{aligned}$$

For $q - 1$ even, a closed-form expression for the integral follows from Lemma A.1 (a). For $q - 1$ odd, note that

$$\int_{-\infty}^{\infty} |\omega|^{q-1} \exp[-\tfrac{1}{2}(\omega + b_w)^2] d\omega = \int_{-\infty}^{\infty} \omega^{q-1} \exp[-\tfrac{1}{2}(\omega + b_w)^2] d\omega - 2 \int_{-\infty}^0 \omega^{q-1} \exp[-\tfrac{1}{2}(\omega + b_w)^2] d\omega$$

so that a closed-form expression can be deduced from Lemma A.1 (a) and (b).

A.3 Determination of Approximate Least Favorable Distributions

A.3.1 Overview

The algorithm is a modified version of what is suggested in Elliott, Müller, and Watson (2015). Let the set $\Theta_c = \{\theta_1, \dots, \theta_m\} \subset \Theta$ be a candidate support for the least favorable measure Λ_c , which is fully characterized by the m nonnegative values λ_j it assigns to $\theta_j \in \Theta_c$. Denote by H_{Λ_c} the corresponding confidence set of the form described in Section A.1.¹⁷ We determine λ_j by an iterative procedure, starting with equal mass on all m points, and then adjusting λ_j as a function of $P_{\theta_j}(g(\theta_j) \notin H_{\Lambda_c}(X, Y))$. In each iteration, λ_j is increased if $P_{\theta_j}(g(\theta_j) \notin H_{\Lambda_c}(X, Y)) > \alpha - \epsilon$ and decreased if $P_{\theta_j}(g(\theta_j) \notin H_{\Lambda_c}(X, Y)) < \alpha - \epsilon$, until numerical convergence to the measure Λ_c^* . The parameter $\epsilon > 0$ induces slight overcoverage of $H_{\Lambda_c^*}$ on Θ_c , so that even with an imperfect candidate Θ_c (and numerically determined Λ_c^*),

¹⁷Here and below, we omit the superscripts ρ , β and σ if the statement applies to all three types of confidence sets.

it is possible that $H_{\Lambda_c^*}$ has coverage uniformly on Θ . This is checked by a numerical search for the maximum of the $\Theta \mapsto [0, 1]$ non-coverage function $RP(\theta) = P_\theta(g(\theta) \notin H_{\Lambda_c^*}(X, Y))$. To this end, it is particularly convenient to employ an importance sampling approximation to $RP(\theta)$, which generates a continuously differentiable approximation, so that standard gradient search algorithms can be employed. If these searches (using random starting points) do not yield a maximum above α , a nearly (up to the parameter $\epsilon > 0$) optimal least favorable measure Λ_c^* has been determined. If the searches yield a θ_0 for which $RP(\theta_0) > \alpha$, then this θ_0 is added to the candidate set Θ_c , and the algorithm iterates.

For the confidence set H^ρ , we seek a family of measures Λ_r^ρ that, for each $r \in (-1, 1)$, have support on the subspace of $\Theta_r = \{\theta : g^\rho(\theta) = r\}$. We discretize this problem into a finite number of values of r . For each given r , we apply the above algorithm, except that the non-coverage function $RP(\theta)$ now only needs to be searched over Θ_r .

We discuss details in the following subsections.

A.3.2 Parameterization

Since the algorithm involves optimization over Θ (or Θ_r), it is convenient to introduce a reparameterization so that this search can be conducted in a unit hypercube. The (A, B, c, d) model is described by 11 parameters. The restriction to invariant sets reduces the number of effective parameters to $11 - 3 = 8$ for H^β and H^σ , and the combination of the bivariate scale invariance and the restriction $\Theta_r = \{\theta : g^\rho(\theta) = r\}$ also makes Θ_r effectively 8 dimensional. The effective parameter space can hence be covered by a $[0, 1]^8 \mapsto \Theta$ function. In particular, given $\eta = (\eta_1, \dots, \eta_8) \in [0, 1]^8$, we set

$$\begin{aligned} c_i &= 2(200)^{2\eta_i-1}, \quad d_i = -0.4 + 1.4\eta_{2+i} \\ r_\eta &= (2\eta_5 - 1) \min(\sqrt{\eta_6\eta_7}, \sqrt{(1-\eta_6)(1-\eta_7)}), \quad \phi_\eta = \pi\eta_8 \\ B &= R \text{chol} \begin{pmatrix} \eta_6 & r_\eta \\ r_\eta & \eta_7 \end{pmatrix}, \quad A = R \text{chol}(I_2 - BB')O(\phi_\eta)S_{c,d} \end{aligned}$$

where $\text{chol}(\cdot)$ is the Choleski decomposition of a matrix, $O(\phi_\eta)$ is the 2×2 rotation matrix for the angle ϕ_η , and $S_{c,d} = \text{diag}(\sqrt{q/\text{tr} \Sigma_X(c_1, d_1)}, \sqrt{q/\text{tr} \Sigma_X(c_2, d_2)})$, with $\Sigma_X(c_0, d_0)$ the $q \times q$ covariance matrix of X in the (A, B, c, d) model when $A = I_2$, $B = 0$, $c_1 = c_0$ and $d_1 = d_0$ (so $\Sigma_X(c_0, d_0)$ is the covariance matrix in the scalar c, d model employed in Müller and Watson (2016) without additional white noise). For H^β and H^σ , we set $R = I_2$. For

H^ρ , we enforce $\theta \in \Theta_r$ by setting $R = \text{chol} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$. The lower and upper bounds for c_1 and c_2 of 0.01 and 400 are such that the distribution of (X, Y) from the resulting $\Sigma_X(c_j, d_0)$ s is nearly indistinguishable from the distribution under the limits $c_j \rightarrow 0$ and $c_j \rightarrow \infty$.

The rationale of this parameterization is that under the equivariance governing H^β and H^σ , it is without loss of generality to consider the case where $\Omega(\theta) = qI_2$. Now both $A = O(\phi_\eta)S_{c,d}$ and $B = 0$, as well as $A = 0$ and $B = I_2$, induce $\Omega(\theta) = qI_2$ with $(2\pi)^{-1}$ as the factor of proportionality for the local-to-zero spectrum $S_z(\omega)$ given in the text. The parameterization of BB' in terms of (η_5, η_6, η_7) exhaustively describes all decompositions of $I_2 = BB' + (I_2 - BB')$ into two positive semidefinite matrices BB' and $(I_2 - BB')$. Under the bivariate scale invariance governing H^ρ , it is without loss of generality to consider the case where $\Omega(\theta) = \begin{pmatrix} 1 & g^\rho(\theta) \\ g^\rho(\theta) & 1 \end{pmatrix}$, and on Θ_r , $g^\rho(\theta) = r$.

A.3.3 Computation of $\Sigma(\theta)$

Gradient methods require fast evaluation of the likelihood for generic θ , which depends on $\Sigma(\theta)$. We initially compute and store the $q \times q$ matrices $\Sigma_X(c_0, d_0)$ introduced in the last subsection for all combinations of the values $c_0 \in \{2(200)^{2i/50-1}\}_{i=0}^{50}$ and $d_0 \in \{-0.4 + 1.4i/40\}_{i=0}^{40}$ using the algorithm developed in Müller and Watson (2016). For a general θ , we then compute $\Sigma_X(c_1, d_1)$ and $\Sigma_X(c_2, d_2)$ by two-dimensional quadratic interpolation of the matrix elements, and construct $\Sigma(\theta)$ via

$$\Sigma(\theta) = (A \otimes I_q) \begin{pmatrix} \Sigma_X(c_1, d_1) & 0 \\ 0 & \Sigma_X(c_2, d_2) \end{pmatrix} (A \otimes I_q)' + (BB' \otimes I_q).$$

A.3.4 Importance Sampling

For H^ρ we employ the importance sampling approximation

$$P_\theta(g^\rho(\theta) \notin H_{\Lambda_c^*}^\rho(X, Y)) \approx N^{-1} \sum_{i=1}^N \frac{f^s(X_{(i)}^s, Y_{(i)}^s | \theta)}{f_p^s(X_{(i)}^s, Y_{(i)}^s)} \mathbf{1}[g^\rho(\theta) \notin H_{\Lambda_c^*}^\rho(X_{(i)}^s, Y_{(i)}^s)] \quad (\text{A.3})$$

for some proposal density f_p^s , where $(X_{(i)}^s, Y_{(i)}^s)$ are i.i.d. draws from f_p^s . For given r , this obviously induces a smooth approximating function on Θ_r , since for all $\theta \in \Theta_r$, $g(\theta) = r$, so that the indicator function does not vary with θ . In fact, for given $H_{\Lambda_c^*}^\rho$, it suffices to compute the sum over those i where $r \notin H_{\Lambda_c^*}^\rho(X_{(i)}^s, Y_{(i)}^s)$, no matter the value of $\theta \in \Theta_r$.

For H^β , note that by equivariance, the event $g^\beta(\theta) \in H_{\Lambda_c^*}^\beta(X, Y)$ is equivalent to $W^\beta = (U_{11}g^\beta(\theta) - U_{12})/U_{22} \in H_{\Lambda_c^*}^\beta(X^\dagger, Y^\dagger)$, where $X^\dagger = (1, X_0^*, X^{*'})'$ and $Y^\dagger = (1, 0, Y^{*'})'$. Thus, given that (X^\dagger, Y^\dagger) are functions of Z^* , we have

$$P_\theta(g^\beta(\theta) \notin H_{\Lambda_c^*}^\beta(X, Y)) \approx N^{-1} \sum_{i=1}^N \frac{f_0^\beta(Z_{(i)}^*, W_{(i)}^\beta | \theta)}{f_p^\beta(Z_{(i)}^*, W_{(i)}^\beta)} \mathbf{1}[W_{(i)}^\beta \notin H_{\Lambda_c^*}^\beta(X_{(i)}^\dagger, Y_{(i)}^\dagger)] \quad (\text{A.4})$$

for some proposal density f_p^β , where $(Z_{(i)}^*, W_{(i)}^\beta)$ are i.i.d. draws from f_p^β . Analogously, for H^σ ,

$$P_\theta(g^\sigma(\theta) \notin H_{\Lambda_c^*}^\sigma(X, Y)) \approx N^{-1} \sum_{i=1}^N \frac{f_0^\sigma(Z_{(i)}^*, W_{(i)}^\sigma | \theta)}{f_p^\sigma(Z_{(i)}^*, W_{(i)}^\sigma)} \mathbf{1}[W_{(i)}^\sigma \notin H_{\Lambda_c^*}^\sigma(X_{(i)}^\dagger, Y_{(i)}^\dagger)] \quad (\text{A.5})$$

with $W^\sigma = g^\sigma(\beta)/|U_{22}|$. These approximation functions are again continuously differentiable in θ , and for given $H_{\Lambda_c^*}^j$, it suffices to perform the summation over those i where $W_{(i)}^j \notin H_{\Lambda_c^*}^j(X_{(i)}^\dagger, Y_{(i)}^\dagger)$, $j \in \{\beta, \sigma\}$.

For the importance sampling approximations to work well, it is crucial that the proposal density f_p is never much smaller than f_0 for all $\theta \in \Theta$ (or never much smaller than f^s over Θ_r in the case of H^ρ). Otherwise, a single large weight $f_0(Z^*, W | \theta)/f_p(Z^*, W)$ may dominate the sums in (A.3), (A.4) and (A.5), inducing imprecise approximations. It is not *a priori* obvious how to construct such a proposal, though, since the densities depend on the fairly high dimensional θ in a complicated way.

To overcome this difficulty, we numerically construct a mixture proposal f_p such that $f_0(z^*, w | \theta)/f_p(z^*, w) \leq \text{IS}_{\max}$ uniformly in θ and (z^*, w) . For simplicity, we describe the approach only in the notation that is relevant for H^β and H^σ :

1. Randomly choose θ_p^1 and set $M = 1$.
2. Using a gradient search algorithm, numerically solve

$$\max_{(z^*, w), \theta} \frac{f_0^j(z^*, w | \theta)}{\sum_{i=1}^M f_0^j(z^*, w | \theta_p^i)}$$

using up to 250 BFGS searches with randomly chosen starting points.

- (a) If the numerically obtained maximum is no larger than IS_{\max} , then set $f_p(z^*, w) = M^{-1} \sum_{i=1}^M f_0^j(z^*, w | \theta_p^i)$ and conclude.

- (b) Otherwise, set θ_p^{M+1} equal to the maximizing value of θ , increase M by one, and iterate step 2.

We set IS_{\max} equal to 2000, 6000 and 12000 for $q \leq 12$, $12 < q \leq 20$ and $q > 20$, respectively.

A.3.5 Computation of Credible Sets and Integrals over F

We approximate integrals over F by a discrete sum over 1000 points θ_j^F , where jointly uniformly distributed random variables are approximated by a low-discrepancy sequence. To ensure that (X, Y) and $(X, -Y)$ have the exact same distribution under our approximation of $\int f(X, Y|\theta)dF(\theta)$, the 1000 points are split into 500 corresponding pairs.

Note that it is not necessary to compute the credible sets H_0 for each realization of $(X_{(i)}^s, Y_{(i)}^s)$ or $Z_{(i)}^*$. Rather, it suffices to determine whether $r \in H_0^p(X_{(i)}^s, Y_{(i)}^s)$ or $W_{(i)} \in H_0(X_{(i)}^\dagger, Y_{(i)}^\dagger)$, respectively. Under the discrete approximation to F , it hence suffices to check whether or not

$$\frac{\sum_{j=1}^{1000} \mathbf{1}[g^p(\theta_j^F) \leq r] f^s(X_{(i)}^s, Y_{(i)}^s | \theta_j^F)}{\sum_{j=1}^{1000} f^s(X_{(i)}^s, Y_{(i)}^s | \theta_j^F)} \quad (\text{A.6})$$

and, for $j \in \{\beta, \sigma\}$,

$$\frac{\int \sum_{j=1}^{1000} \mathbf{1}[w \leq W_{(i)}^j] f_0(Z_{(i)}^*, w | \theta_j^F) dw}{\sum_{j=1}^{1000} f_1^*(Z_{(i)}^* | \theta_j^F)} \quad (\text{A.7})$$

take on values in the interval $[\alpha/2, 1 - \alpha/2]$, respectively. We compute the integral in (A.7) by numerical quadrature.

Since all three type of confidence sets always contain H_0 , the realizations of $(X_{(i)}^s, Y_{(i)}^s)$ and $(Z_{(i)}^*, W_{(i)}^j)$ for which (A.6) and (A.7) take on values between $[\alpha/2, 1 - \alpha/2]$ never enter the sums (A.3), (A.4) and (A.5) that approximate the non-rejection probabilities. The effective number of terms in the sums is thus greatly reduced, which correspondingly facilitates computations. With this in mind, we modify the determination of the importance sampling proposal by maximizing the (empirical analogue of the) variance of the importance sampling weights conditional on the event $g(\theta) \notin H_0$.

A.3.6 Approximate Least Favorable Distributions and Size Control

The initial candidate set Θ_c consists of 10 randomly selected points in Θ (or in Θ_r in the case of H^r). For given Θ_c , Λ_c^* is computed by the algorithm described in Elliott, Müller, and Watson (2015), using a target value the level of $1 - \alpha + \epsilon$. We set ϵ to 0.3%, 0.6% and 1.0% for $\alpha = 5\%$, 10% and 33% for $q \leq 12$, respectively, double these values for $12 < q \leq 20$, and triple them for $q > 20$. We search for coverage violating points by BFGS maximizations over the importance sampling approximation to the non-coverage probability function $RP(\theta)$, using numerical derivatives and random starting values. We collect up to 5 coverage violating points in up to 50 BFGS searches before augmenting Θ_c and recomputing Λ_c^* , which is fairly time consuming, especially if Θ_c consists of many points. The algorithm stops once 500 consecutive BFGS searches do not yield a coverage violating point.

A.3.7 Quality of Approximation and Time to Compute

With $N = 250,000$ importance sampling draws and the baseline case of $q = 12$, the Monte Carlo standard errors of non-coverage probabilities are approximately 0.1%-0.25% at the 5% level, 0.1%-0.35% at the 10% level, and 0.3%-0.5% at the 33% level. Using results in Elliott, Müller, and Watson (2015) and Müller and Watson (2016), it is straightforward to use the approximately least favorable distributions to obtain lower bounds on the F -weighted average expected length of any confidence set of nominal level. We find that our sets are within approximately 1-3% of this lower bound, so they come reasonably close to being as short as possible under that criterion.

For $q = 12$, a specific level α and problem, the determination of the approximately least favorable measure Λ^* takes approximately 5 minutes using a Fortran implementation on a dual 10-core PC, and yields an approximate least favorable measure Λ^* with approximately 30-100 points of support. Running times are roughly quadratic in q due the $4q^2$ elements in the quadratic forms of the likelihoods. Larger q also lead to bigger Monte Carlo standard errors of rejection probabilities, as the importance sampling now must cover an effectively larger set of distributions of (X^s, Y^s) and (Z^*, W) , respectively.

A.4 Data Used

The data and sources are listed in Table A.1.

A.5 Long-run projections and low-pass filters

Figure A.1 plots growth rates of GDP and consumption growth rates along with low-pass moving averages designed to isolate variation in the series with periods longer than 11 years and the long-run projections using $q = 12$ shown in Figure 1 of the paper. The low-pass moving averages were computed using an ideal low-pass filter for periods longer than $T/6$ (≈ 11 years) truncated after $T/2$ terms. The series were padded with pre- and post-sample backcasts and forecasts constructed from an AR(4) model. The long-run projection were computed as the projections of GDP growth rates onto Ψ_T with $q = 12$, including constant term.

A.6 Long-run covariances using a subset of the columns of Ψ_T

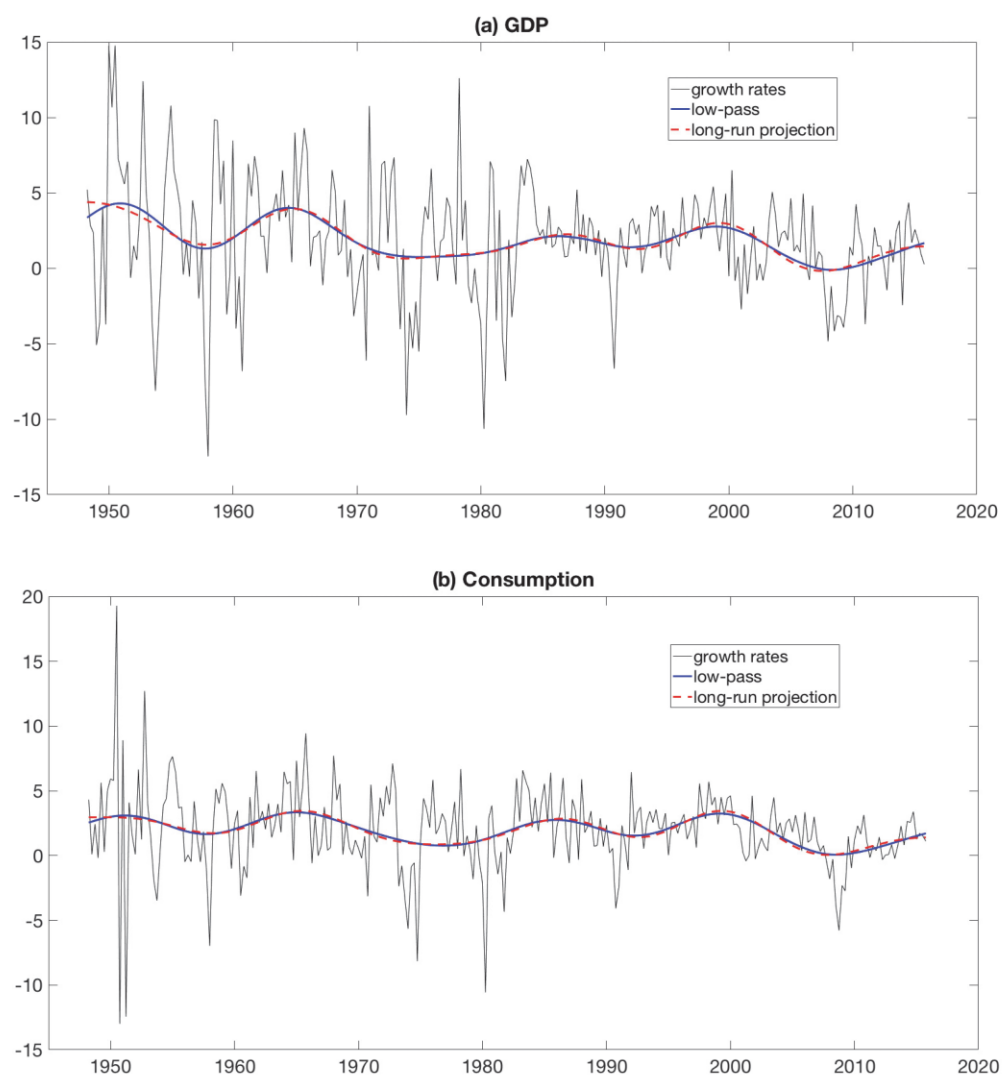
The empirical results in the body of the paper rely on covariance measures associated with projections of the data onto q cosine functions capturing periodicities of between $2T$ and $2T/q$, where T is the length of the sample. Using data from 1948-2015 ($T = 68$ years) this analysis used periods longer than 11 years to define “long-run” variation and covariation, so $q = 12$. While 11 years is longer than typical business cycles, it does incorporate periods corresponding to what some researchers refer to as the “medium run” (Blanchard (1997), Comin and Gertler (2006)). In this appendix we consider measures of long-run covariability that focus on a subset of the q periods. This allows a comparison of, say, results from periods corresponding to the “medium-long run” and to those from the “longer-long run.”

To motivate the new measures, look at Figure 1 in the text which plots the projections of GDP and consumption growth rates onto $q = 12$ cosine regressors with periods that range from $T/6$ (≈ 11 years) to $2T$ (136 years). Figure A.2 show the corresponding projections onto the first $q_1 = 6$ of these cosine terms (with periods from $T/3 \approx 23$ years to $2T = 136$ years) and last $q_2 = 6$ cosine terms (with periods $T/12 \approx 11$ years to $2T/7 \approx 19$ years). The

Table A.1: Data series used

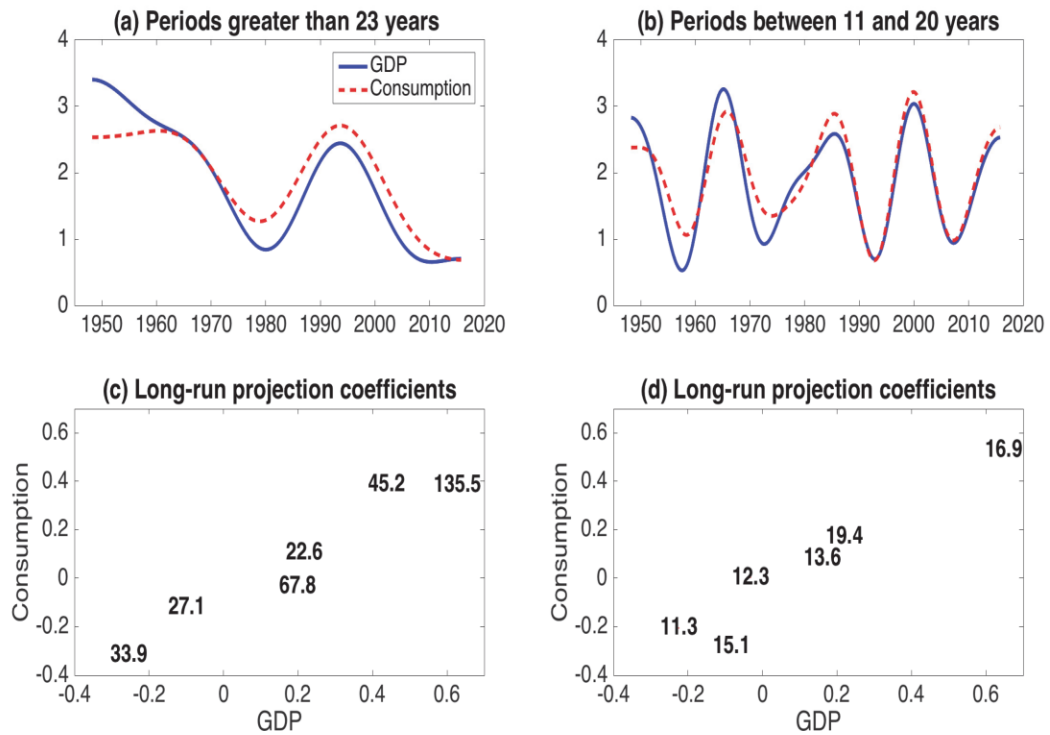
Series	Sources and Notes (FRED Codes)
GDP, consumption, investment, and employee compensation	NIPA nominal values (GDP, PCDG, PCND, PCESV, GDPI, PNFI, PRFI, Y033RC1Q027SBEA, COE) deflated by the price index for personal consumer expenditures (PCECTPI). The variables are expressed in per-capita terms using the $q = 12$ low-frequency projection of civilian non-institutionalized population (CNP16OV).
TFP	Growth rate for TFP from Fernald (2014), updated from his webpage.
Interest rates	3-Month Treasury bill rate (TB3MS) and 10-Year Treasury bond rate (GS10)
Inflation	Inflation from the personal consumption deflator (PCECTPI) and consumer price index (CPIAUCSL)
Money supply	M1 money supply (M1) from FRB beginning in 1959:M1. This is linked to M1 (currency + demand deposits) from Friedman and Schwartz (1963, Table A-1, Col. 7)
Unemployment rate	Bureau of Labor Statistics (UNRATE)
Stock returns	CRSP Nominal Monthly Returns are from WRDS. Monthly real returns were computed by subtracting the change in the logarithm in the CPI from the nominal returns, which were then compounded to yield quarterly returns. Values are $400 \times$ the logarithm of gross quarterly real returns.
Stock prices, dividends, and earnings	S&P composite prices, dividends, and earnings from Robert Shiller's webpage (file IE.XLS).
Exchange rates and relative CPIs	Nominal exchange rate (EXUSUK) from the FRB, CPI for the UK from the Bank of England (CPIUKQ) and U.S. CPI (CPIAUCSL) from the BLS.

Figure A.1: Long-run average growth rates of GDP and consumption



Notes: The first show growth rates, low-pass moving averages (periods 11 years and greater) and projections onto $q = 12$ cosine terms.

Figure A2: Long-run projections for GDP and consumption growth rates for different periodicities



Notes: Panel (a) plots the projections of the data onto six cosine terms with periods 23-136 years. Panel (b) shows the projections onto six low-frequency terms with periods 11-19 years. Sample means have been added to both sets of projections. Panels (c) and (d) are scatterplots of the coefficients (cosine transforms) from panels (a) and (b) where the plot symbols are the periods (in years) of the associated cosine function.

first of these captures the longer-long-run variation in the data, and the second captures the medium-long-run variability. Each can be studied separately. To differentiate these periodicities, we replace equation (4) with

$$\Omega_{q_1:q_2,T} = T^{-1} \sum_{t=1}^T E \left[\begin{pmatrix} \hat{x}_{q_1:q_2,t} \\ \hat{y}_{q_1:q_2,t} \end{pmatrix} \begin{pmatrix} \hat{x}_{q_1:q_2,t} & \hat{y}_{q_1:q_2,t} \end{pmatrix} \right] = E \begin{bmatrix} X'_{q_1:q_2,T} X_{q_1:q_2,T} & X'_{q_1:q_2,T} Y_{q_1:q_2,T} \\ Y'_{q_1:q_2,T} X_{q_1:q_2,T} & Y'_{q_1:q_2,T} Y_{q_1:q_2,T} \end{bmatrix} \quad (\text{A.8})$$

where the subscript “ $q_1 : q_2$ ” notes that the projection is computed using the q_1 through q_2 cosine terms (i.e., the q_1 through q_2 columns of Ψ_T) corresponding to periods $2T/q_2$ through $2T/q_1$. Thus the longer-long-run periodicities shown in Figure A.2.a correspond to the covariance matrix $\Omega_{1:6,T}$ (the first 6 cosine terms) and the medium-long-run periodicities in Figure A.2.b correspond to $\Omega_{7:12,T}$ (the 7-12th cosine terms).

Throughout the paper we have used q to denote the number of low-frequency cosine terms that define the long-run periods of interest (perhaps divided further into longer-long and medium-long). But q plays another important role in the analysis. The value of Ω (or now $\Omega_{q_1:q_2}$) ultimately depends on the variability and persistence in the stochastic process as exhibited in the local-to-zero (pseudo-) spectrum S_z . This spectrum is parameterized by (A, B, c, d) ; see equation (11). We learn about the value of these parameters (and therefore the value of Ω) using the data $(X_{1:q,T}, Y_{1:q,T})$. Thus, q also denotes the sample variability in the data that is used to infer the value of the long-run covariance matrix Ω . So, while our interest might lie in the longer-long-run covariability captured in $\Omega_{1:6}$, the sample variability in $(X_{1:12,T}, Y_{1:12,T})$ might be used to learn about $\Omega_{1:6}$. While it is arguably most natural to match the variability in the data used for inference to the variability of interest, for example using $(X_{1:q,T}, Y_{1:q,T})$ to learn about $\Omega_{1:q}$, if the (A, B, c, d) model accurately characterizes the spectrum over a wider frequency band, then variability over this wider band can improve inference. But of course using a wider frequency band runs the risk of misspecification if the (A, B, c, d) model is a poor characterization of the spectrum over this wider range of frequencies. This is the standard trade-off of robustness and efficiency.

With these ideas in mind, Table A.2 shows results for long-run correlation and regression parameters from $\Omega_{1:12}$, $\Omega_{1:6}$, and $\Omega_{7:12}$, corresponding the periods $T/6$ and higher, $T/3$ and higher, and $T/6$ through $2T/7$. Results are shown using inference based on the same $q = 12$ cosine transforms used in the body of the paper, but also using $q = 6$, so only lower frequency variability in the data is used to learn about (A, B, c, d) , and with $q = 18$, so higher frequency variability is also used. (For simplicity, we use these values of q for all series, even though

the sample period is shorter for long-term interest rates and exchange rates.)

The first block of results in Table A.2 are for consumption and GDP. The first row repeats earlier results using the $q = 12$ cosine terms to learn about $\Omega_{q_1:q_2}$ with $q_1 = 1$ and $q_2 = 12$. The other rows are for other values of q , q_1 , and q_2 . The results suggest remarkable stability across the different values of q , q_1 , and q_2 . Figure A.2.c and A.2.d provides hints at this stability. It shows the scatter plot of $(X_{1:6,T}, Y_{1:6,T})$ and $(X_{7:12,T}, Y_{7:12,T})$ corresponding to the projections plotted in panels (a) and (b). The scatter plots corresponding to the different periodicities are quite similar, and this is reflected in the stability of the results shown in Table A.2. This same stability across q , q_1 , and q_2 is evident for many of the other pairs of variables. Looking closely at Table A.2, there are subtle differences in the rows. For example, the confidence intervals for the parameters from $\Omega_{1:12}$ tend to be somewhat narrower using $q = 18$ than using $q = 12$, consistent with a modest amount of additional information using a larger value of q . The same result holds for results for $\Omega_{1:6}$ computed using $q = 6$ and $q = 12$.

Table A.2: Long-run covariation measures for selected variables with $\Omega_{q_1:q_2}$
Inference based on q cosine transforms.

Y	X	q	q1:q2	$\rho_{q_1:q_2}$			$\hat{\beta}$	$\beta_{q_1:q_2}$		$\hat{\sigma}_{q(q_2, \beta)}$
				$\hat{\rho}$	67% CI	90% CI		67% CI	90% CI	
Cons.	GDP	12	1:12	0.91	0.83, 0.96	0.71, 0.97	0.77	0.66, 0.87	0.48, 0.95	0.41
		12	1:6	0.92	0.82, 0.96	0.60, 0.98	0.74	0.57, 0.87	0.36, 0.97	0.31
		12	7:12	0.92	0.84, 0.96	0.74, 0.98	0.80	0.69, 0.92	0.60, 1.04	0.23
	18	1:12	0.91	0.84, 0.95	0.76, 0.97	0.75	0.66, 0.84	0.52, 0.91	0.40	
	6	1:6	0.89	0.71, 0.95	0.50, 0.98	0.73	0.57, 0.89	0.36, 1.08	0.31	
Inv.	GDP	12	1:12	0.53	0.29, 0.72	0.02, 0.81	1.24	0.64, 1.79	0.21, 2.21	2.18
		12	1:6	0.54	0.27, 0.76	-0.02, 0.91	1.18	0.58, 1.79	-0.46, 2.27	1.54
		12	7:12	0.55	0.25, 0.73	0.04, 0.82	1.36	0.70, 2.03	-0.15, 2.88	1.36
	18	1:12	0.63	0.42, 0.75	0.29, 0.81	1.60	1.05, 2.16	0.61, 2.59	2.39	
	6	1:6	0.56	0.21, 0.79	-0.04, 0.91	1.07	0.52, 1.67	-0.14, 2.28	1.13	
$w \times n$	GDP	12	1:12	0.98	0.96, 0.99	0.94, 0.99	1.28	1.20, 1.36	1.14, 1.42	0.31
		12	1:6	0.98	0.96, 0.99	0.93, 0.99	1.28	1.19, 1.36	1.12, 1.42	0.21
		12	7:12	0.97	0.95, 0.98	0.89, 0.99	1.29	1.20, 1.39	1.08, 1.54	0.22
	18	1:12	0.95	0.91, 0.97	0.87, 0.98	1.21	1.12, 1.31	1.04, 1.38	0.46	
	6	1:6	0.97	0.92, 0.98	0.83, 0.99	1.26	1.14, 1.38	1.04, 1.48	0.23	
GDP	TFP	12	1:12	0.78	0.64, 0.89	0.45, 0.95	1.22	0.92, 1.48	0.72, 1.72	0.74
		12	1:6	0.79	0.63, 0.94	0.44, 0.97	1.22	0.92, 1.54	0.69, 1.78	0.54
		12	7:12	0.75	0.58, 0.86	0.38, 0.91	1.16	0.84, 1.48	0.60, 1.72	0.44
	18	1:12	0.76	0.67, 0.87	0.50, 0.94	1.21	0.97, 1.48	0.76, 1.69	0.75	
	6	1:6	0.71	0.41, 0.94	0.08, 0.97	1.14	0.71, 1.53	0.32, 2.00	0.57	

Table A.2: continued

Y	X	q	$q_1:q_2$	$\rho_{q_1:q_2}$			$\beta_{q_1:q_2}$			$\hat{\sigma}_{q_1:q_2, X}$
				$\hat{\rho}$	67% CI	90% CI	$\hat{\beta}$	67% CI	90% CI	
Cons. (Nondur)	GDP	12	1:12	0.41	0.08, 0.61	-0.09, 0.72	0.35	0.13, 0.57	-0.07, 0.76	0.88
		12	1:6	0.42	0.07, 0.62	-0.54, 0.86	0.32	0.08, 0.57	-0.41, 0.76	0.64
		12	7:12	0.42	0.08, 0.65	-0.14, 0.73	0.40	0.13, 0.64	-0.07, 0.96	0.56
		18	1:12	0.41	0.14, 0.56	-0.20, 0.76	0.33	0.15, 0.51	-0.17, 0.65	0.83
		6	1:6	0.32	-0.03, 0.60	-0.30, 0.86	0.37	-0.03, 0.74	-0.37, 1.11	0.76
Cons. (Serv.)	GDP	12	1:12	0.84	0.71, 0.90	0.57, 0.96	0.83	0.66, 0.99	0.54, 1.24	0.60
		12	1:6	0.87	0.74, 0.95	0.58, 0.98	0.88	0.70, 1.18	0.54, 1.45	0.47
		12	7:12	0.82	0.67, 0.88	0.45, 0.94	0.70	0.51, 0.87	0.35, 0.99	0.32
		18	1:12	0.84	0.72, 0.92	0.63, 0.96	0.81	0.66, 0.97	0.54, 1.25	0.62
		6	1:6	0.83	0.60, 0.95	0.34, 0.98	0.93	0.66, 1.22	0.42, 1.51	0.47
Cons. (Dur.)	GDP	12	1:12	0.78	0.68, 0.90	0.53, 0.92	1.86	1.46, 2.25	1.15, 2.56	1.52
		12	1:6	0.82	0.67, 0.93	0.48, 0.96	1.72	1.33, 2.12	0.93, 2.47	1.05
		12	7:12	0.77	0.67, 0.89	0.56, 0.94	2.12	1.63, 2.78	1.28, 3.36	1.02
		18	1:12	0.75	0.63, 0.83	0.52, 0.91	1.80	1.41, 2.20	1.11, 2.50	1.82
		6	1:6	0.87	0.70, 0.94	0.47, 0.98	1.45	1.13, 1.84	0.83, 2.23	0.67
Inv. (Nonres.)	GDP	12	1:12	0.42	0.13, 0.65	-0.05, 0.75	0.97	0.41, 1.46	-0.09, 1.90	2.18
		12	1:6	0.47	0.15, 0.72	-0.09, 0.89	0.97	0.41, 1.59	-0.09, 2.02	1.51
		12	7:12	0.38	0.07, 0.62	-0.20, 0.73	0.90	0.22, 1.52	-0.59, 2.14	1.47
		18	1:12	0.55	0.38, 0.73	0.14, 0.78	1.32	0.81, 1.78	0.36, 2.18	2.23
		6	1:6	0.53	0.20, 0.79	-0.05, 0.94	0.91	0.43, 1.40	-0.16, 1.94	1.00

Table A.2: continued

Y	X	q	$q_1:q_2$	$\rho_{q_1:q_2}$			$\hat{\rho}$	$\beta_{q_1:q_2}$			$\hat{\sigma}_{q_1:q_2,y x}$
				67% CI	90% CI			$\hat{\beta}$	67% CI	90% CI	
Inv. (Res.)	GDP	12	1:12	0.40	0.10, 0.64	-0.10, 0.72		2.15	0.77, 3.53	-0.27, 4.56	5.63
		12	1:6	0.44	0.08, 0.63	-0.16, 0.86		1.97	0.60, 3.35	-0.61, 4.56	3.67
		12	7:12	0.40	0.07, 0.65	-0.12, 0.72		2.32	0.77, 4.56	-1.30, 6.80	4.08
		18	1:12	0.46	0.29, 0.63	0.07, 0.74		2.49	1.27, 3.86	0.35, 4.78	5.68
		6	1:6	0.45	0.09, 0.72	-0.13, 0.84		1.61	0.54, 2.78	-0.31, 4.06	2.18
Inv. (Equip.)	GDP	12	1:12	0.33	0.00, 0.54	-0.20, 0.70		0.81	0.12, 1.57	-0.41, 2.11	2.75
		12	1:6	0.36	-0.00, 0.58	-0.38, 0.83		0.89	0.05, 1.65	-0.64, 2.18	1.99
		12	7:12	0.29	-0.00, 0.56	-0.20, 0.70		0.89	0.05, 1.73	-0.87, 2.56	1.76
		18	1:12	0.50	0.29, 0.69	0.07, 0.75		1.33	0.61, 2.04	-0.10, 2.53	2.88
		6	1:6	0.32	-0.03, 0.64	-0.46, 0.77		0.89	-0.07, 1.94	-1.21, 2.99	1.95
10Y nom. rates	3M nom. rates	12	1:12	0.96	0.92, 0.98	0.89, 0.98		0.92	0.83, 1.06	0.75, 1.15	0.70
		12	1:6	0.96	0.91, 0.98	0.88, 0.98		0.93	0.83, 1.09	0.75, 1.18	0.61
		12	7:12	0.94	0.88, 0.97	0.76, 0.98		0.87	0.67, 0.97	0.60, 1.05	0.28
		18	1:12	0.95	0.91, 0.98	0.83, 0.98		0.87	0.75, 1.02	0.67, 1.12	0.80
		6	1:6	0.96	0.92, 0.98	0.83, 0.99		0.97	0.86, 1.09	0.75, 1.20	0.61
10Y real rates	3M real rates	12	1:12	0.95	0.89, 0.96	0.80, 0.98		0.98	0.86, 1.09	0.71, 1.26	0.68
		12	1:6	0.95	0.87, 0.97	0.72, 0.98		0.98	0.85, 1.17	0.65, 1.34	0.57
		12	7:12	0.95	0.88, 0.97	0.83, 0.98		0.97	0.85, 1.08	0.76, 1.17	0.31
		18	1:12	0.89	0.79, 0.94	0.71, 0.98		0.91	0.77, 1.06	0.66, 1.24	0.83
		6	1:6	0.94	0.86, 0.97	0.70, 0.98		1.02	0.87, 1.17	0.64, 1.42	0.61

Table A.2: continued

Y	X	q	q1:q2	ρ_{q_1,q_2}			$\hat{\beta}$	β_{q_1,q_2}		$\hat{\sigma}_{q_1,q_2,Y^X}$
				$\hat{\rho}$	67% CI	90% CI		67% CI	90% CI	
CPI Infl.	PCE Infl.	12	1:12	0.98	0.96, 0.99	0.95, 0.99	1.13	1.07, 1.19	0.98, 1.24	0.36
		12	1:6	0.98	0.95, 0.99	0.94, 0.99	1.12	1.06, 1.18	0.96, 1.24	0.29
		12	7:12	0.98	0.95, 0.99	0.92, 0.99	1.17	1.10, 1.33	1.04, 1.41	0.17
		18	1:12	0.97	0.95, 0.98	0.94, 0.99	1.14	1.08, 1.20	0.99, 1.26	0.37
		6	1:6	0.98	0.96, 0.99	0.93, 0.99	1.11	1.04, 1.17	0.97, 1.22	0.28
3M rates	PCE Infl.	12	1:12	0.47	0.21, 0.83	-0.00, 0.91	0.73	0.34, 1.49	-0.09, 1.91	2.20
		12	1:6	0.52	0.24, 0.89	-0.02, 0.95	0.84	0.38, 1.59	0.05, 2.02	1.92
		12	7:12	0.23	-0.05, 0.56	-0.38, 0.70	0.36	-0.14, 0.72	-0.64, 1.07	0.88
		18	1:12	0.54	0.28, 0.83	0.00, 0.91	0.82	0.47, 1.47	0.03, 1.91	2.20
		6	1:6	0.65	0.30, 0.89	0.00, 0.96	1.02	0.55, 1.59	0.13, 2.20	1.98
10Y rates	PCE Infl.	12	1:12	0.47	0.23, 0.83	-0.00, 0.91	0.66	0.30, 1.39	-0.06, 1.73	2.12
		12	1:6	0.52	0.24, 0.89	-0.02, 0.94	0.73	0.33, 1.43	0.03, 1.86	1.89
		12	7:12	0.23	0.03, 0.58	-0.29, 0.71	0.33	0.00, 0.70	-0.36, 0.96	0.79
		18	1:12	0.49	0.22, 0.83	0.03, 0.91	0.64	0.29, 1.28	0.08, 1.63	2.03
		6	1:6	0.57	0.23, 0.89	-0.03, 0.95	0.88	0.47, 1.43	0.01, 2.02	2.02
CPI Infl.	Mon. Supply	12	1:12	0.12	-0.17, 0.54	-0.60, 0.76	0.11	-0.17, 0.52	-0.58, 0.90	2.45
		12	1:6	0.16	-0.15, 0.54	-0.60, 0.83	0.13	-0.17, 0.52	-0.63, 1.04	2.15
		12	7:12	0.01	-0.35, 0.31	-0.66, 0.47	-0.00	-0.33, 0.30	-0.77, 0.54	0.96
		18	1:12	0.13	-0.13, 0.42	-0.36, 0.76	0.11	-0.10, 0.56	-0.28, 0.91	2.24
		6	1:6	0.27	-0.08, 0.59	-0.37, 0.86	0.27	-0.08, 0.66	-0.39, 1.09	2.34

Table A.2: continued

Y	X	q	q1:q2	ρ_{q_1,q_2}			$\hat{\beta}$	β_{q_1,q_2}		$\hat{\sigma}_{q_1,q_2,Y}$
				$\hat{\rho}$	67% CI	90% CI		67% CI	90% CI	
Un. Rate	PCE Infl.	12	1:12	0.26	-0.03, 0.60	-0.27, 0.80	0.21	-0.04, 0.45	-0.24, 0.76	1.44
		12	1:6	0.27	-0.03, 0.55	-0.28, 0.83	0.21	-0.04, 0.47	-0.26, 0.83	1.24
		12	7:12	0.09	-0.19, 0.43	-0.43, 0.56	0.12	-0.17, 0.43	-0.63, 0.67	0.66
		18	1:12	0.21	-0.04, 0.54	-0.46, 0.80	0.18	-0.06, 0.42	-0.24, 0.78	1.47
		6	1:6	0.27	-0.08, 0.57	-0.38, 0.86	0.23	-0.07, 0.49	-0.36, 0.85	1.34
Un. Rate	TFP	12	1:12	-0.65	-0.75, -0.34	-0.91, -0.13	-1.00	-1.39, -0.62	-1.65, -0.27	1.06
		12	1:6	-0.78	-0.92, -0.47	-0.97, -0.22	-1.22	-1.61, -0.79	-1.96, -0.44	0.75
		12	7:12	-0.31	-0.56, 0.10	-0.69, 0.38	-0.53	-1.09, 0.21	-1.39, 0.64	0.66
		18	1:12	-0.42	-0.76, -0.26	-0.91, -0.08	-0.92	-1.31, -0.49	-1.70, -0.15	1.17
		6	1:6	0.27	-0.08, 0.57	-0.38, 0.86	0.23	-0.07, 0.49	-0.36, 0.85	1.34
3M real rates	Consumption	12	1:12	0.42	0.08, 0.60	-0.06, 0.80	0.88	0.32, 1.45	-0.11, 2.57	1.84
		12	1:6	0.42	0.06, 0.64	-0.38, 0.80	0.88	0.26, 1.57	-1.11, 3.26	1.48
		12	7:12	0.43	0.14, 0.68	-0.13, 0.76	0.82	0.32, 1.38	-0.11, 1.76	0.88
		18	1:12	0.35	0.03, 0.54	-0.13, 0.80	0.63	0.11, 1.16	-0.28, 2.41	1.76
		6	1:6	0.30	-0.05, 0.59	-0.46, 0.86	0.97	-0.19, 2.03	-1.48, 3.43	1.71
Stock returns	Consumption	12	1:12	0.40	0.07, 0.60	-0.08, 0.80	2.85	0.98, 4.94	-0.48, 8.48	6.15
		12	1:6	0.43	0.08, 0.65	-0.16, 0.86	3.27	0.98, 5.57	-0.90, 10.78	4.86
		12	7:12	0.35	0.04, 0.62	-0.16, 0.73	2.44	0.56, 4.31	-0.90, 5.77	3.17
		18	1:12	0.42	0.16, 0.60	-0.02, 0.76	3.05	1.47, 4.80	0.24, 7.43	6.12
		6	1:6	0.41	0.00, 0.70	-0.23, 0.89	4.23	0.62, 7.84	-2.63, 11.46	5.65

Table A.2: continued

Y	X	q	$q_1:q_2$	ρ_{q_1,q_2}		90% CI	β_{q_1,q_2}			$\hat{\sigma}_{q_1,q_2} K$
				$\hat{\rho}$	67% CI		$\hat{\beta}$	67% CI	90% CI	
Stock prices	Dividends	12	1:12	0.20	-0.05, 0.43	-0.30, 0.72	0.45	-0.17, 1.06	-0.60, 1.68	7.27
		12	1:6	0.36	-0.00, 0.66	-0.25, 0.91	0.76	-0.05, 1.56	-0.60, 3.16	5.26
		12	7:12	0.03	-0.46, 0.29	-0.72, 0.46	-0.11	-0.72, 0.69	-1.22, 1.31	4.03
		18	1:12	0.39	0.14, 0.55	-0.06, 0.72	0.75	0.31, 1.24	-0.08, 1.63	6.34
		6	1:6	0.42	0.02, 0.71	-0.20, 0.89	1.15	0.28, 1.94	-0.50, 2.72	5.56
Stock prices	Earnings	12	1:12	0.21	-0.04, 0.42	-0.27, 0.57	0.38	-0.15, 0.92	-0.53, 1.35	7.23
		12	1:6	0.29	-0.06, 0.52	-0.29, 0.76	0.49	-0.10, 1.19	-0.64, 2.16	5.23
		12	7:12	0.11	-0.15, 0.35	-0.46, 0.50	0.17	-0.26, 0.65	-0.64, 1.14	4.40
		18	1:12	0.38	0.09, 0.44	-0.09, 0.56	0.24	0.05, 0.44	-0.10, 0.60	6.62
		6	1:6	0.23	-0.08, 0.57	-0.37, 0.73	0.64	-0.22, 1.50	-0.99, 2.18	5.92
Exchange rates	Rel. price ind.	12	1:12	0.42	0.13, 0.57	-0.06, 0.72	1.19	0.51, 1.95	0.00, 2.54	6.10
		12	1:6	0.42	0.11, 0.56	-0.16, 0.80	0.93	0.26, 1.61	-0.33, 2.29	4.10
		12	7:12	0.51	0.14, 0.67	0.03, 0.80	1.87	0.85, 3.43	0.17, 4.70	4.10
		18	1:12	0.38	0.16, 0.54	-0.06, 0.63	1.01	0.43, 1.68	-0.02, 2.19	5.57
		6	1:6	0.41	0.00, 0.70	-0.27, 0.91	0.54	0.04, 0.95	-0.38, 1.29	2.08

Notes: Results are based on Ω_{q_1,q_2} (col. 4 lists q_1 and q_2) and sample information in q cosine transforms (col. 3 shows q).

Additional References

DHRYMES, P. J. (2005): “Moments of Truncated (Normal) Distributions,” *Working Paper, Columbia University*.

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. Wiley, New York.

FERNALD, J. (2014): “A Quarterly, Utilization-Adjusted Series on Total Factor Productivity,” *FRBSF Working Paper 2012-9*.

FRIEDMAN, M., AND A. J. SCHWARTZ (1963): *A Monetary History of the United States, 1867-1960*. Princeton University Press, Princeton NJ.