Long-Run Covariability[*]

Ulrich K. Müller and Mark W. Watson

Princeton University

Department of Economics

Princeton, NJ, 08544

First draft: August 2016
This version: January 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 trends computed from the data. These trends are similar to low-pass filtered data and are designed to extract variability corresponding to periods longer than the span of the sample divided by $q/2$, where q is a small number, such as 12. 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 the "great ratios", nominal exchange rates and relative nominal prices, unemployment rate and inflation, money growth and inflation, earnings and stock prices, etc.

JEL classification: C22, C53, E17

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

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. First, 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. The second problem is that there are few "long-run" observations in the samples typically used in empirical analyses of long-run relations, so sample information is limited. 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.

The most well-known example of faulty inference due to a mistaken assumption about persistence is Granger and Newbold's (1974) "spurious regression", where standard OLS inference leads to grossly misleading conclusions when applied to independent $I(1)$ variables. The last 40 years has seen important progress developing inference for specific classes of stochastic processes (most notably for $I(0)$ and integrated/cointegrated processes), but several aspects of the resulting inference remains fragile. For example, while HAC standard errors lead to reliable inference in $I(0)$ settings with limited serial correlation, the resulting hypothesis tests exhibit substantial size distortions for stationary series with high serial correlation (e.g., den Haan and Levin (1997), Kiefer, Vogelsang, and Bunzel (2000), and Müller (2014)). Inference in cointegrated models is well-developed (e.g., Engle and Granger (1987), Johansen (1988), Phillips (1991), Stock and Watson (1993)), but these models have knife-edge implications about long-run covariability (cointegrated variables have unit long-run correlations) and efficient inference methods are not robust to small departures from the model's assumed exact unit autoregressive roots (Elliott (1998)). Variables that are highly but not perfectly correlated in the long-run, or are highly persistent, but perhaps without exact unit roots,

fall outside the standard cointegration framework.

This paper develops methods designed to provide reliable inference about long-run co-variability for a wide range of persistence patterns (encompassing $I(0)$, $I(1)$, and many other forms of long-run persistence) and that are applicable regardless of the degree of long-run correlation. The methods rely on low-frequency averages of the data to measure the data's long-run variability and covariability. These long-run data summaries have proven useful for constructing long-run covariance matrices and associated test statistics in $I(0)$ settings (e.g., Müller (2004, 2007), Phillips (2005), Sun (2013), and Lazarus, Lewis, and Stock (2016)), but also for conducting inference about more general patterns of long-run persistence and measuring uncertainty about long-run predictions (Müller and Watson (2008, 2016)). A key simplification offered by these averages is that they are normally distributed in large samples even though the stochastic process generating the data may exhibit substantial persistence (Müller and Watson (forthcoming)). This allows large-sample inference about covariability parameters to be transformed into a finite-sample problem involving a handful of normal random variables and, while the inference problem is "non-standard," it can be solved using previously developed statistical methods paired with modern computing power.

The paper's goal is to provide empirical researchers with an 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 both valid over a wide range of persistence patterns and nearly optimal in the sense of having close to shortest expected length (see Section 4 for details). As discussed in Section 3, the procedures allow 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 defines the notion of long-run variability and covariability used throughout the paper. These are defined in terms of 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)), Hodrick and Prescott (1997)). The discussion is carried out in the context of two empirical examples, the long-

---

[1]The replication files contains a matlab function for computing these confidence intervals, available at www.princeton.edu\~mwatson. The function uses the approximate least favorable distributions discussed in Section 4 and the appendix, which are also available in the replication files.

run relationship between consumption and GDP and between short- and long-term nominal interest rates. In the long-run projections we employ, long-run variability and covariability is equivalently captured by the covariability of a small number of trigonometrically weighted averages of the data. Section 3 derives the large-sample normality of these 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 low dimensional multivariate normal random vector. Section 4 reviews relevant methods for solving this finite sample problem. Section 5 uses the resulting inference methods 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. Section 6 examines the robustness of Section 5's empirical conclusions to changes in the periodicities defining the "long-run", and to alternative choices for the information set used for inference.

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 will 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 consumption/income ratio is one of economics' "Great Ratios" (Klein and Kosobud (1961)); the dynamic implications of this stability inspired early work on error-correction models (e.g., Sargan (1964) and Davidson, Hendry, Srba, and Yeo (1978)), and these in turn motivated Granger's formulation of cointegration (Granger (1981)). While early analysis provided empirical support for the cointegration of consumption and income (e.g., Campbell (1987), King, Plosser, Stock, and Watson (1991), Cochrane (1994)), more recent work has come

to the opposite conclusion (see Lettau and Ludvigson (2013) for discussion and references). Whether or not consumption and income are cointegrated (i.e., have an exact unit autoregressive root and exact unit long-run correlation), even a casual glance at the data suggests the two variables move together closely in the long run.
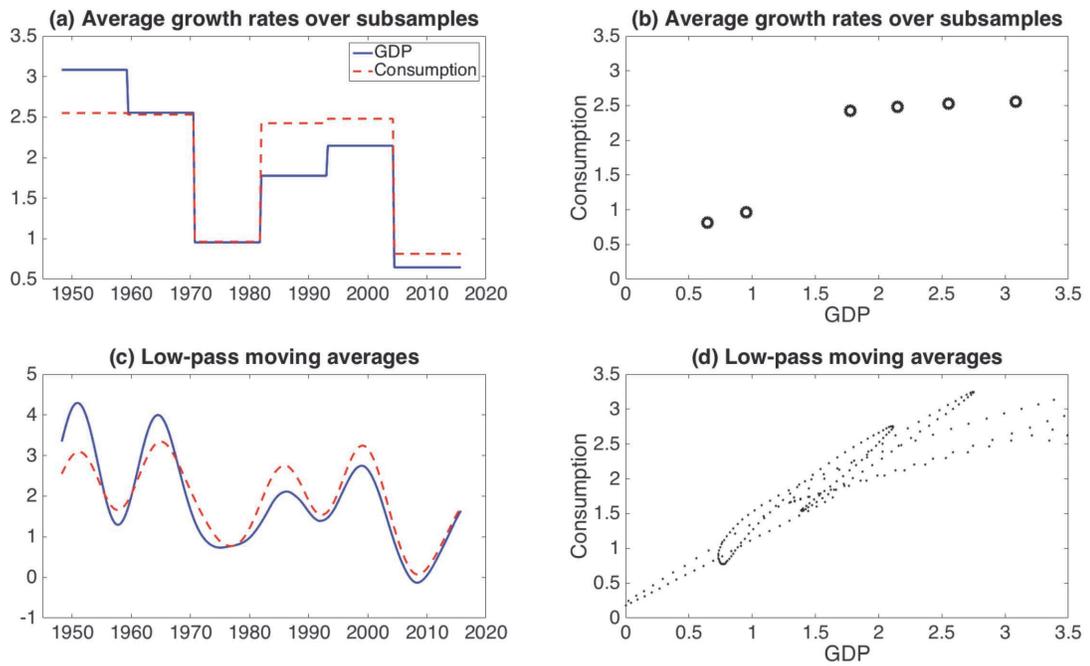
Consider, for example, the evolution of U.S. real per-capita GDP and consumption over the post-WWII period. In the 17 years from 1948 through 1964, GDP increased by 62% and consumption increased by 52%. Over the next 17 years (1965-1981) both GDP and consumption grew more slowly, by only 30%. Growth 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%). In this sense, GDP and consumption exhibited substantial long-run variability and covariability over the post-WWII period.[2]

There are two distinct notions of "long-run" implicit in this calculation. The most obvious is that each period makes up 17 years, approximately twice the length of the typical business cycle. But another is that each period encompasses a large fraction (1/4) of the full 1948-2015 sample period. Our statistical framework defines long-run in this latter way: long-run statistical analysis involves inference about characteristics of stochastic processes that govern the evolution of averages of the data over periods that are large relative to the available sample.

With this is 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 averages growth rates against time, and Figure 1.b is a scatterplot of the six average growth rates for consumption against corresponding values for GDP. Each of the six sub-samples contains 11.25 years (45 quarters), spans of history longer than the typical business cycle, and arguably capture "long-run" variability in GDP and consumption. And, each represents

---

[2]Consumption 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 GDP and consumption

(a) Average growth rates over subsamples

(b) Average growth rates over subsamples

(c) Low-pass moving averages

(d) Low-pass moving averages

Notes: Panel (a) shows sample averages of the variables over the period shown. Panel (c) shows low-pass moving averages using a frequency cutoff corresponding to $T/6$ periods (approximately 11 years). Panels (b) and (d) are scatterplots of the variables in (a) and (c).

a substantial fraction (1/6) of the sample and is a long-run observation in the statistical sense discussed in the preceding paragraph. Average GDP and consumption growth over these subsamples exhibited substantial variability and (from the scatter plot) roughly one-for-one covariability.

Figure 1.c sharpens the analysis by plotting "low-pass" moving averages of the series designed to isolate variation in the series with periods longer than 11 years.[3] Sample variation in these moving averages is much like the variation in the subsample averages of Figure 1.a, but Figure 1.c captures the smooth transition of the series from high-growth to low-growth periods. The scatterplot of these moving averages is plotted in Figure 1.d. Like Figure 1.b, it shows the close relationship between long-run movements in consumption and GDP, but it also shows the high degree of serial correlation in the moving averages.

A convenient device for handling this serial correlation is to use projections on low-frequency periodic functions in place of the low-pass moving averages. To be specific, let x_t, $t = 1, ..., T$ 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), $\Psi(s) = [\Psi_1(s), \Psi_2(s), \ldots, \Psi_q(s)]'$ 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 $2T/j$. Most of our empirical analysis uses $q = 12$ which captures periodicities longer than $T/6$; this defines the long-run variation in the data the analysis is designed to capture. The projection of x_t onto $\Psi((t-1/2)/T)$ for $t = 1, ..., T$ yields the fitted values

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

where X_T are the projection (linear regression) coefficients, $X_T = (\Psi_T'\Psi_T)^{-1}\Psi_T'x_{1:T}$, where $x_{1:T}$ is the $T \times 1$ vector with t'th element given by x_t. The matrix Ψ_T has two properties that simplify calculations and interpretation. First, $\Psi_T'l_T = 0$ where l_T is a vector of ones, so that $\widehat{x}_t$ also corresponds to the projection of $x_t - \overline{x}_{1:T}$ onto $\Psi((t-1/2)/T)$, where $\overline{x}_{1:T}$ is the sample mean. Second, $T^{-1}\Psi_T'\Psi_T = I_q$, so X_T corresponds to simple cosine-weighted

---

[3]These were computed using an ideal low-pass filter for periods longer than $T/6$ truncated after $T/2$ terms. The series were padded with pre- and post-sample backcasts and forecasts constructed from an AR(4) model.

Figure 2: Long-run projections of GDP and consumption growth rates
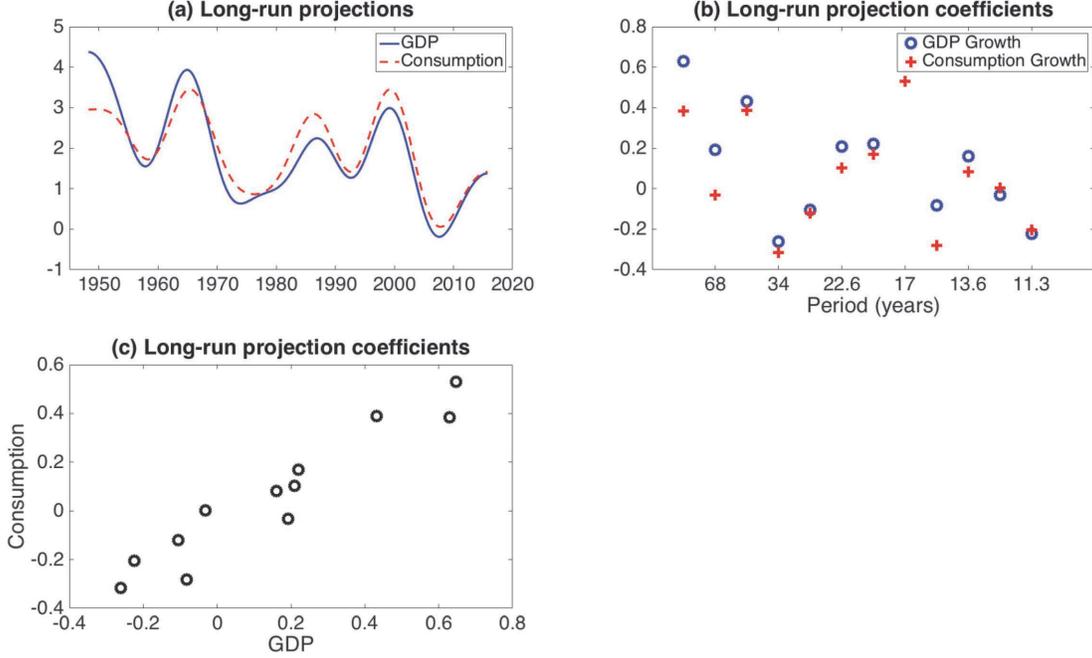
Notes: Panel (a) plots the projections of the data onto the low-frequency cosine terms discussed in the text, where sample means have been added to projections so they are consistent with the low-pass moving averages plotted in Figure 1(c). Panel (b) plots the projection coefficients (X_{jT}, Y_{jT}) against period $2T/j$ (in years). Panel (c) is a scatterplot of the variables from (b).

averages of the data (i.e., are the "cosine transforms" of $\{x_t\}$)

$$X_T = T^{-1} \sum_{t=1}^{T} \Psi((t-1/2)/T)x_t. \tag{2}$$

Letting (x_t, y_t) denote the growth rates of GDP and consumption, the long-run projections $(\widehat{x}_t, \widehat{y}_t)$ are plotted in Figure 2.a. Except for minor differences near the endpoints, these long-run projections essentially coincide with the low-pass moving average plotted in Figure 1.c, so both capture the same long-run sample variability in the data. An advantage of the long-run projections is that they are fully summarized by the projection coefficients (X_T, Y_T), a relatively small number of cosine-weighted averages of the sample data. Figure 2.b plots the projection coefficients, (X_{jT}, Y_{jT}) against the period of the corresponding co-

7

sine term, $2T/j$. Evidently, there is substantial variation and covariation in the projection coefficients. Indeed, the scatterplot of (X_{jT}, Y_{jT}) shown in Figure 2.c suggests a roughly one-to-one relationship between the cosine transforms.

The orthogonality of the cosine regressors Ψ_T leads to a tight connection between the variability and covariability in the long-run projections $(\widehat{x}_t, \widehat{y}_t)$ plotted in Figure 2.a and the cosine transforms (X_{jT}, Y_{jT}) plotted in Figure 2.b and 2.c:

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

Thus, sample covariability in the time series projections coincides with sample covariability in the cosine transforms.

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 bills, x_t, and the rate on 10-year U.S. Treasury bonds, y_t, from 1953 through 2015. The levels of these interest rates are highly serially correlated, but the term spread, $y_t - x_t$, far less so. Early cointegration work (e.g., Campbell and Shiller (1987)) modeled the level of interest rates as $I(1)$, and short- and long-rates as cointegrated. Later empirical analysis of the term structure (e.g., Dai and Singleton (2000), Diebold and Li (2006)) model the levels of interest rates as a function of small number of dynamic common factors that lead to common, but less than unit-root, long-run persistence.

Figure 3 plots the levels of short- and long-term interest rates, (x_t, y_t), along with their long-run projections, $(\widehat{x}_t, \widehat{y}_t)$, and cosine transforms, (X_T, Y_T). The long-run projections capture the rise in interest rates from the beginning of the sample through the early 1980s and then their subsequent decline (Figure 3.a). These long-swings in the level of interest rates lead to relatively larger values in the long-period cosine transforms (Figure 3.b). The projections for long-term interest rates closely track the projections for short-term rates and, given the connection between between the projections and cosine transforms, X_{jT} and Y_{jT} are highly correlated (Figure 3.c).

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 $(\widehat{x}_t, \widehat{y}_t)$ or equivalently (from (3)) the projection coefficients (X_T, Y_T). This suggests that the covariance/variance

8

Figure 3: Short- and long-term interest rates
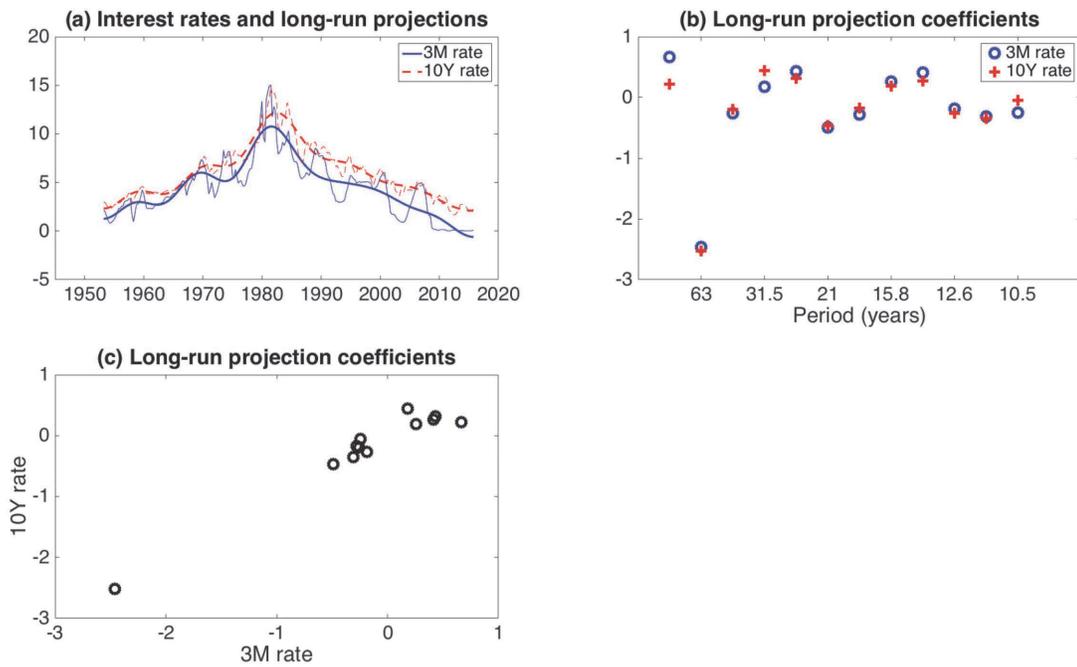
Notes: Panel (a) plots the data and projections of the data onto the low-frequency cosine terms discussed in the text, where sample means have been added to projections). Panel (b) plots the projection coefficients (X_{jT}, Y_{jT}) against period $2T/j$ (in years). Panel (c) is a scatterplot of the variables from (b).

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

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

where the equalities directly follow from (3).

The 2×2 matrix Ω_T is the average covariance matrix of the long-run projections $(\widehat{x}_t, \widehat{y}_t)$ 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

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

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

$$
\beta_T = \arg \min_b E\left[T^{-1} \sum_{t=1}^{T} (\widehat{y}_t - b\widehat{x}_t)^2\right],
$$

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 $\hat{x}_t$,[4] $\sigma^2_{y|x,T}$ is the average variance of the prediction error, and $\rho^2_{xy,T}$ is the corresponding population R^2. These parameters thus measure the population

---

[4]The parameter β_T is closely related to a linear band-spectrum regression coefficient (Engle (1974)), corresponding to periods longer than $2T/q$.

comovement of the long-run variation of (x_t, y_t). Equivalently, by the second equality in (4), β_T also solves

$$\beta_T = \arg\min_b E\left[\sum_{j=1}^{q}(Y_{jT} - bX_{jT})^2\right]$$

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

The objective of the remaining analysis is to develop inference about the parameters $(\rho_{xy,T}, \beta_T, \sigma^2_{y|x,T})$.

3 Asymptotic approximations and parameterizing long-run persistence and covariability

The long-run correlation and regression parameters 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), which 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, and thus for the parameters of interest $\rho_{xy,T}$, β_T, and $\sigma^2_{y|x,T}$. The second issue involves parameterizing the form of long-run persistence and comovement which 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 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, let $z_t = (x_t, y_t)'$ and 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, the coefficients in $C_T(L)$ die out sufficiently fast that 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

11

1).[5] If the spectral density converges for all frequencies close to zero

$$T^{3-2\kappa} F_{\Delta z, T}(\omega/T) \to S_{\Delta z}(\omega)$$

in a suitable sense, then

$$T^{1-\kappa} \begin{pmatrix} X_T \\ Y_T \end{pmatrix} \Rightarrow \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \Sigma), \tag{6}$$

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

$$T^{2-2\kappa} \operatorname{Var} \begin{pmatrix} X_T \\ Y_T \end{pmatrix} = T^{2-2\kappa} \Sigma_T \to \Sigma. \tag{7}$$

The limiting covariance matrix Σ in (6) and (7) is a function of the "local-to-zero" spectrum $S_{\Delta z}$ and the cosine weights $\Psi_j(s)$ that determine (X_T, Y_T); see Müller and Watson (forthcoming) for additional details and an explicit formula. We make three comments about these large-sample results. First, they hold when the *first-difference* of z_t has a spectral density (and therefore has limited persistence); 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 sample mean of z_t also has a large-sample normal limit. Second, in common parameterizations of persistence models, the scale factor $T^{-\kappa}$ depends on the form of persistence; for example, the factor is $T^{-1/2}$ for $I(0)$ persistence and $T^{-3/2}$ for $I(1)$ persistence. However, we focus on inference procedures that do not depend on the scale of z_t (due to invariance or equivariance), so $T^{-\kappa}$ does not need to be known. Third, because $T^{2-2\kappa} \Sigma_T \to \Sigma$, then $T^{2-2\kappa} \Omega_T \to \Omega$ where Ω is defined as in the last expression of (4) with Σ in place of Σ_T. Correspondingly, $(\rho_{xy,T}, \beta_T, T^{2-2\kappa} \sigma^2_{y|x,T}) \to (\rho_{xy}, \beta, \sigma^2_{y|x})$ with the limits defined by (5) with Ω in place of Ω_T. Thus, a solution to the small-sample problem of inference about $(\rho_{xy}, \beta, \sigma^2_{y|x})$ from observing (X, Y) readily translates into a large-sample solution to inference about $(\rho_{xy,T}, \beta_T, \sigma^2_{y|x,T})$.

---

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

3.2 Parameterizing long-run persistence and covariability

The limiting 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 for the first-difference of z, $S_{\Delta z}$. The corresponding local-to-zero (pseudo-) spectrum for the level of z_t is $S_z(\omega) = \omega^{-2} S_{\Delta z}(\omega)$. In this section we discuss parameterizations of S_z, 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 $S_z(\omega) \propto \Lambda$, and straightforward calculations show that $\Sigma = \Lambda \otimes I_q$ and $\Omega \propto \Lambda$, 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 2 and 3 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 $(\rho_{xy}, \beta, \sigma^2_{y|x})$ 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 2 and 3 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, consumption, short-, and long-term interest rates: Table 1 presents estimates and confidence sets for $(\rho_T, \beta_T, \sigma^2_{y|x.T})$ using (X_T, Y_T) with $q = 12$ for GDP and consumption (panel a) and short- and long-term interest rates (panel b). Results are presented for $I(0)$ and $I(1)$ models, and for a more general model of persistence introduced below. For now, focus on the $I(0)$ and $I(1)$ results. The point estimates shown in the table are MLEs, and confidence intervals for $(\beta_T, \sigma^2_{y|x.T})$ 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 $I(1)$

Table 1: Long-run covariability estimates and confidence intervals from the $I(0)$, $I(1)$, and (A,B,c,d) models

a. GDP and consumption

		ρ	β	$\sigma_{y\mid x}$
$I(0)$	Estimate	0.93	0.76	0.35
	67% CI	0.87, 0.96	0.67, 0.85	0.30, 0.46
	90% CI	0.81, 0.97	0.60, 0.92	0.26, 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
(A,B,c,d)	Estimate	0.91	0.77	0.41
	67% CI	0.83, 0.96	0.66, 0.87	0.32, 0.54
	90% CI	0.71, 0.97	0.48, 0.95	0.28, 0.67
	67% Bayes CS	*0.83, 0.96*	*0.66, 0.87*	*0.32, 0.54*
	90% Bayes CS	*0.71, 0.97*	*0.58, 0.95*	*0.28, 0.67*

b. Short- and long-term interest rates

		ρ	β	$\sigma_{y\mid x}$
$I(0)$	Estimate	0.97	0.96	0.63
	67% CI	0.95, 0.98	0.89, 1.03	0.53, 0.81
	90% CI	0.93, 0.99	0.84, 1.08	0.47, 0.97
$I(1)$	Estimate	0.94	0.85	0.48
	67% CI	0.88, 0.96	0.76, 0.95	0.40, 0.62
	90% CI	0.82, 0.97	0.68, 1.03	0.36, 0.74
(A,B,c,d)	Estimate	0.96	0.92	0.70
	67% CI	0.92, 0.98	0.83, 1.06	0.53, 0.92
	90% CI	0.89, 0.98	0.75, 1.15	0.47, 1.36
	67% Bayes CS	*0.92, 0.97*	*0.83, 1.00*	*0.53, 0.92*
	90% Bayes CS	*0.89, 0.98*	*0.75, 1.07*	*0.47, 1.13*

Notes: 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 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.

estimates and confidence sets. The estimated long-run correlation is greater than 0.9, 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 — for example, there is strong evidence that the series are highly correlated over the long-run — although the $I(0)$ and $I(1)$ results differ more sharply than for GDP and consumption. To take just one example, the 90% confidence interval for ρ ranges from 0.82 to 0.97 in the $I(1)$ model but is narrower (0.93 to 0.99) in the $I(0)$ model.

As we show in Table 2 below, the $I(0)$ assumption yield 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. This raises the question of how to obtain valid inference in both models, and, more generally, under a wider range of forms of persistence.

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. Thus, parameterizing S_z is a crucial issue for inference about long-run covariability. 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 constrained to be tractable. $I(0)$ 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 tractable, 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

$$S_z(\omega) \propto A \begin{pmatrix} (\omega^2 + c_1^2)^{-d_1} & 0 \\ 0 & (\omega^2 + c_2^2)^{-d_2} \end{pmatrix} A' + BB' \tag{8}$$

where A is unrestricted and B is lower triangular.[6]

This model generates the standard spectral shapes: $A = 0$ yields the $I(0)$ model; $B = 0$, $c = 0$ and $d_1 = d_2 = 1$ yields the $I(1)$ model; $B = 0$, $d_1 = d_2 = 1$ yields a model with two AR roots local-to-unity; $B = 0$ and $c = 0$ yields a bivariate fractional model. Other choices of (A, B, c, d) yield models that combine persistent and non-persistent components (as in cointegrated or "local-level" models) but go beyond the usual $I(0)/I(1)$ or fractional formulations. The cost of the (A, B, c, d) model's flexibility is that it contains 11 parameters as opposed to just 3 in the $I(0)$ and $I(1)$ models. Yet, as we discuss in the next section, it is still possible conduct valid inference even in the loosely parameterized (A, B, c, d) model.

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 $\sigma_{y|x}$. As a general matter, the goal is to compute confidence intervals that are as informative ("narrow") 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. ($\gamma = \rho$, β, or $\sigma_{y|x}$ for the problem we consider). Let $H(X, Y)$ denote a confidence interval for γ and $\text{vol}(H(X, Y))$ denote the length of the interval. The objective is to choose H so that it has small expected length, $E[\text{vol}(H(X, Y))]$, 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

---

[6]This 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.

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. As a general matter, let W denote a distribution that puts weight on different values of θ, so the problem becomes

$$\min_{H} \int E_{\theta}(\text{vol}(H(X, Y))) dW(\theta) \tag{9}$$

subject to

$$\sup_{\theta \in \Theta} P_{\theta}\left(\gamma \in H(X, Y)\right) \geq 1 - \alpha \tag{10}$$

where the objective function (9) emphasizes that the expected volume depends on the value of θ, with different values of θ weighted by W, and the coverage constraint (10) 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 $H_0 : \gamma = \gamma_0$. The solution to (9)-(10) 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 (close to) smallest weighted expected length, 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 $1 - \alpha$ Bayes credible sets.[7]

---

[7]Numerical calculations show that the Müller and Norets (2016) adjustment has a small (3%-8%) effect on expected length of 95%, 90%, and 67% confidence intervals for all three parameters of interest.

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:

$$\rho \in H^\rho(X,Y) \Leftrightarrow \rho \in H^\rho(b_x X, b_y Y) \text{ for } b_x b_y > 0 \tag{11}$$

$$\beta \in H^\beta(X,Y) \Leftrightarrow \frac{b_y \beta + b_{yx}}{b_x} \in H^\beta(b_x X, b_y Y + b_{yx} X) \text{ for } b_x, b_y \neq 0 \text{ and all values of } b_{yx} \tag{12}$$

$$\sigma_{y|x} \in H^\sigma(X,Y) \Leftrightarrow |b_y|\sigma_{y|x} \in H^\sigma(b_x X, b_y Y + b_{yx} X) \text{ for } b_x, b_y \neq 0 \text{ and all values of } b_{yx}. \tag{13}$$

These invariance/equivariance restrictions lead to two modifications to the solution to (9)-(10). 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 (9) 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

The parameter space for $\theta = (A, B, c, d)$ is as follows: 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 $i = 1, 2$.[8] 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 (as, for example, in a cointegrated model).

The confidence sets we construct require three distributions over θ: the weighting function W for computing the average length in the objective (9), the Bayes prior associated

---

[8]See Appendix 3.2 for details.

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 (9)-(10) and is approximated using numerical methods similar to those discussed in Elliott, Müller, and Watson (2015), with details provided in the supplementary appendix. We use the same distribution for W and the Bayes prior. Specifically, the 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 distributed $U[0, 1]$.

4.2.3 Empirical results for GDP, consumption, and interest rates

Table 1 in Section 3.2 above 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. For most entries, the Bayes credible sets are slightly larger than the $I(0)$ sets, presumably reflecting the possibility of persistence greater than $I(0)$, as was evident in Figure 4. 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 are somewhat different. The confidence intervals (and Bayes credible sets) are roughly in-between the $I(0)$ and $I(1)$ intervals. Substantively, the results indicate that long-run movements in short- and long-rates are highly correlated, and that a unit long-run response of long-rates to short-rates is consistent with these data.

Table 2: Coverage rates for efficient 90% confidence intervals with data generated by different stochastic processes

Efficient confidence set for	Data generated by:				
	$I(0)$	$I(1)$	$I(0)+I(1)$	$I(d)$	(A,B,c,d)
$I(0)$	0.90	0.01	0.01	0.01	0.01
$I(1)$	0.00	0.90	0.00	0.00	0.00
$I(0)+I(1)$	0.91	0.91	0.90	0.68	0.68
$I(d)$	0.90	0.90	0.87	0.90	0.87
(A,B,c,d)	0.91	0.90	0.90	0.90	0.90

4.3 Coverage properties of restricted versions of the (A, B, c, d) model

In this subsection we investigate the coverage distortions for confidence intervals constructed using misspecified models of persistence. Specifically we consider five models of persistence, and for each model we both generate data and construct confidence intervals for ρ. The data are generated using $\rho = 0$ and Table 2 shows the fraction of the confidence sets that include the true value $\rho = 0$.[9] The models considered are 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 (8). Because data were generated by and confidence intervals constructed for each of these five models, the table contains 25 entries. The columns indicate the model used to generate the data, the rows shows the model used to construct the confidence set, and the entries are fraction of confidence sets that contain the true value of ρ, minimized over the other parameters used to generate the data. The diagonal entries of the table are 0.90 indicating that each method has coverage 90% under its assumed data generating process. The off-diagonal differ from 0.90 and show the coverage distortions. For example, 90% $I(0)$ confidence sets have coverage of just 1% when the data are generated by the other four models. $I(1)$ confidence sets have similarly bad coverage when the data are not generated by the $I(1)$ model. The $I(0) + I(1)$ model encompasses both the $I(0)$ and $I(1)$ models, so the associated confidence

---

[9]Results are shown for confidence sets that do not incorporate the Müller-Norets Bayes superset adjustment. Including this adjustment yields similar results.

intervals has good coverage for these models, but has coverage of only 68% in the $I(d)$ and (A, B, c, d) models. The $I(d)$ model encompasses the $I(0)$ and $I(1)$ models, and so has good coverage for these models. It does not encompass the the $I(0) + I(1)$ or (A, B, c, d) models, but exhibits only a small coverage distortion in these cases. Finally, the general (A, B, c, d) model encompasses all of the other models, and so controls coverage uniformly across these models.

Table 2 highlights the large coverage distortions associated with confidence intervals based on $I(0)$, $I(1)$, or $I(0) + I(1)$ models. These results echo results in the earlier literature on the fragility of $I(0)$ and $I(1)$ inference (e.g., den Haan and Levin (1997) for HAC inference in $I(0)$ models and Elliott (1998) for inference in cointegrated models). Table 2 suggests that inference based on the $I(d)$ model is much less fragile; indeed it offers near nominal coverage in Table 2. However, the $I(d)$ model does not fare as well in other contexts; for example Müller and Watson (2016) show that $I(d)$ model yields long-run prediction sets with significant undercoverage when data are generated by a univariate analogue of the (A, B, c, d) model.

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 is true for several of the relationships that we now investigate.

Figure 4: Long-run projections for GDP, consumption, investment, labor income
and TFP growth rates
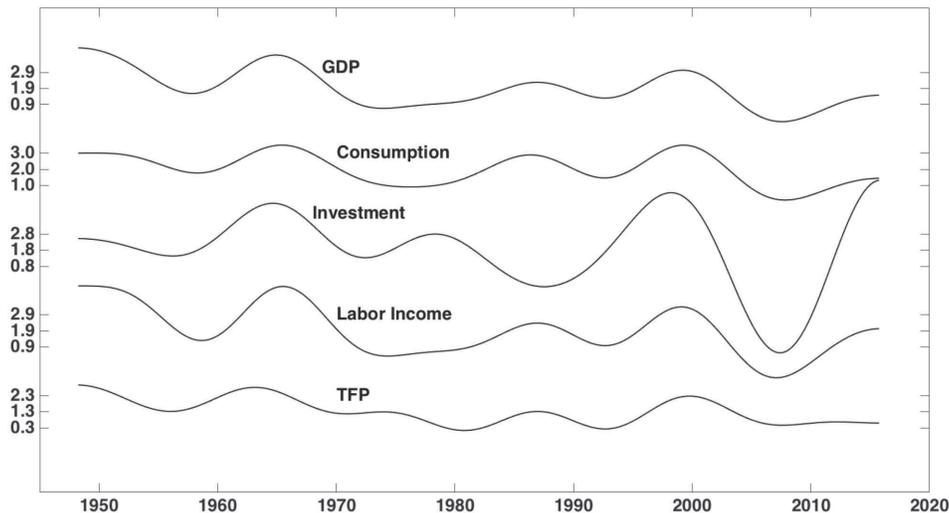
5.1 Balanced growth correlations

In the standard one-sector growth model, variations in per-capita GDP, consumption, invest-
ment, 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 consump-
tion onto the logarithm of income has a unit coefficient, as do the same regressions with
consumption replaced by investment or labor income. A long-run one-percentage point in-
crease 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 evolu-
tion 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.

22

Table 3: Long-run correlations of GDP, consumption, investment,
labor compensation, and TFP

	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)	(0.60, 0.97)	(0.06, 0.82)		0.71 (0.53, 0.84)
TFP	(0.46, 0.95)	(0.29, 0.89)	(-0.08, 0.72)	(0.36, 0.92)	

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.

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 2.a.) The figure indicates substantial long-run covariability over the post-WWII period, but less so for investment than the other variables. Table 3 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 using the general (A, B, c, d) model (shown in parentheses). The values below the main diagonal are the corresponding 90% confidence intervals using the (A, B, c, d) model. Table 4 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 4 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 vari-

Table 4: Selected long-run regressions involving GDP, consumption, investment, labor compensation, and TFP

Y	X		β				$\hat{\sigma}_{y\|x}$
			$\hat{\beta}$	67% CI	90% CI		
Consumption	GDP		0.77	0.66, 0.87	0.48, 0.95		0.41
Investment	GDP		1.24	0.64, 1.79	0.21, 2.21		2.18
Labor comp. ($w \times n$)	GDP		1.28	1.20, 1.36	1.14, 1.42		0.31
GDP	TFP		1.22	0.92, 1.48	0.72, 1.72		0.74
Cons. (Nondurable)	GDP		0.35	0.13, 0.57	-0.07, 0.76		0.88
Cons. (Services)	GDP		0.83	0.66, 0.99	0.54, 1.24		0.60
Cons. (Durables)	GDP		1.86	1.46, 2.25	1.15, 2.56		1.52
Inv. (Nonresidential)	GDP		0.97	0.41, 1.46	-0.09, 1.90		2.18
Inv. (Residential)	GDP		2.15	0.77, 3.53	-0.27, 4.56		5.63
Inv. (Equipment)	GDP		0.81	0.12, 1.57	-0.41, 2.11		2.75

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.

ations 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 (labor income and GDP, consumption and GDP), and second that the long-run regression coefficient on GDP is different from unity for some 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 much larger than 0.8.

5.2   Other long-run relations

Figure 5 and Table 5 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 measured 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

Figure 5: Long-run projections and projection coefficients

(a) PCE and CPI inflation rates

(b) Inflation and 3-month Treasury bill rates

(c) Inflation and 10-year Treasury bond rates

(d) Real 3-Month and 10-Year interest rates

(e) Money supply (M1) growth rate and inflation

(f) Inflation and unemployment rates

Figure 5 (continued)

(g) TFP growth rate and unemployment rate

(h) Consumption growth rate and real 3-month interest rate

(i) Consumption growth rate and real stock returns

(j) Dividend and stock price growth rates

(k) Earnings and stock price growth rates

(l) Relative CPI indices and exchange rates

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.

27

Table 5: Long-run covariation measures for selected variables

	Y	X	Trans.	ρ			β			$\hat{\sigma}_{y|x}$
				$\hat{\rho}$	67% CI	90% CI	$\hat{\beta}$	67% CI	90% CI	
1	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
2	3M rates	PCE Infl.	L,L	0.47	0.21, 0.83	-0.00, 0.91	0.73	0.34, 1.49	-0.09, 1.91	2.20
3	10Y rates	PCE Infl.	L,L	0.47	0.23, 0.83	-0.00, 0.91	0.66	0.30, 1.39	-0.06, 1.73	2.12
4	10Y real rates	3M real rates	L,L	0.95	0.89, 0.96	0.80, 0.98	0.98	0.87, 1.09	0.71, 1.26	0.68
5	CPI Inflation	Money Supply	L,G	0.12	-0.17, 0.54	-0.60, 0.76	0.11	-0.17, 0.52	-0.58, 0.90	2.45
6	Un. Rate	PCE Infl.	L,L	0.26	-0.03, 0.60	-0.27, 0.80	0.21	-0.04, 0.45	-0.24, 0.76	1.44
7	Un. Rate	TFP	L,G	-0.65	-0.75, -0.34	-0.91, -0.13	-1.00	-1.39, -0.62	-1.65, -0.27	1.06
8	3M real rates	Consumption	L,G	0.42	0.08, 0.60	-0.06, 0.80	0.88	0.32, 1.45	-0.11, 2.57	1.84
9	Stock returns	Consumption	L,G	0.40	0.07, 0.60	-0.08, 0.80	2.85	0.98, 4.94	-0.48, 8.48	6.15
10	Stock prices	Dividends	G,G	0.20	-0.05, 0.43	-0.30, 0.72	0.45	-0.17, 1.06	-0.60, 1.68	7.27
11	Stock prices	Earnings	G,G	0.21	-0.04, 0.42	-0.27, 0.57	0.38	-0.15, 0.92	-0.53, 1.35	7.23
12	Exchange rates	Rel. price ind.	G,G	0.42	0.13, 0.57	-0.06, 0.72	1.19	0.51, 1.95	0.00, 2.54	6.10

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.

in PCE and CPI inflation track each other closely and the 90% confidence interval shown in Table 5 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 and table show the long-run covariation of inflation and short- and long-term nominal interest rates. 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.[10] 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: $0.80 \leq \rho \leq 0.98$) 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

---

[10]These 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 (King and Watson (1994)).

from -0.17 to 0.54.

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. 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 5 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 and $I(0)$ 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 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.[11] While there is considerable uncertainty about the value of the long-run correlation between 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 0.72 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). 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 5, the growth rate of nominal exchange rates and relative nominal prices are positively correlated over the long-run, statistically significantly so at the 33% significance level, but the correlation is far from perfect ($\rho < 0.72$ based on the 90% confidence set). We highlight two caveats. First, we use the post-Bretton Woods sample period, so the sample includes only 45 years, and using $q = 12$ cosine terms the long-run projections capture variability with periods of (approximately) 7 years or higher. This 7-year period may be sufficiently short that long-run adjustments have not occurred, something we investigate in the next section. Second, the price ratio uses relative CPIs, a large component of which includes non-traded goods which may be less tightly linked to exchange rates than prices of traded goods.

6   Alternative measures of long-run covariability

The empirical results in the last section relied on covariance measures associated with projections of the data onto $q = 12$ cosine functions capturing periodicities of $T/6$ or higher, 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. While 11

---

[11]The data are for the S&P, and are updated updated versions of the data used in Shiller (2000) available on Robert Shiller's webpage.

Figure 6: Long-run projections for GDP and consumption growth rates for different periodicities
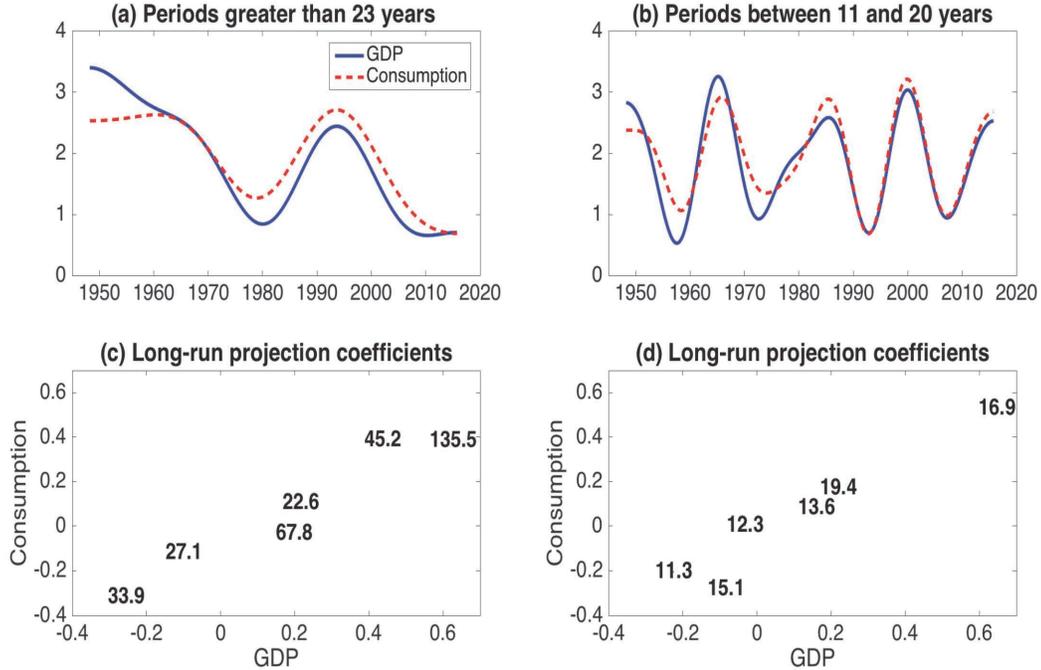
Notes: Panel (a) plots the projections of the data onto six cosine terms with periods 23-136 years. Panel (b) shows the projections onto six low-frequency terms with periods 11-19 years. Sample means have been added to both sets of projections. Panels (c) and (d) are scatterplots of the coefficients (cosine transforms) from panels (a) and (b) where the plot symbols are the periods (in years) of the associated cosine function.

years is longer than typical business cycles, it does incorporates periods corresponding to what some researchers refer to as the "medium run" (Blanchard (1997), Comin and Gertler (2006)). In this section 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 again at Figure 2.a 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). Figures 6.b and 6.c 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 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

$$\Omega_{i:j,T} = T^{-1} \sum_{t=1}^{T} E\left[\begin{pmatrix} \widehat{x}_{i:j,t} \\ \widehat{y}_{i:j,t} \end{pmatrix} \begin{pmatrix} \widehat{x}_{i:j,t} & \widehat{y}_{i:j,t} \end{pmatrix}\right] = E\begin{bmatrix} X'_{i:j,T}X_{i:j,T} & X'_{i:j,T}Y_{i:j,T} \\ Y'_{i:j,T}X_{i:j,T} & Y'_{i:j,T}Y_{i:j,T} \end{bmatrix} \quad (14)$$

where the subscript "$i : j$" notes that the projection is computed using the i through j cosine terms (i.e., the i through j columns of Ψ_T) corresponding to periods $2T/j$ through $2T/i$. Thus the longer-long-run periodicities shown in Figure 6.a correspond to the covariance matrix $\Omega_{1:6,T}$ (the first 6 cosine terms) and the medium-long-run periodicities in Figure 6.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 $\Omega_{i:j}$) 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 (8). 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 6 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 sections above, 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. Table 6.a shows results for long-run covariability of GDP, consumption, investment, labor compensation, and TFP. Table 6.b shows results for selected long-run

Table 6: Long-run covariation measures for selected variables with Ω_{ij}.
Inference based on q cosine transforms.

a. GDP, consumption, investment, labor compensation, and TFP

Y	X	q	$i{:}j$	$\hat{\rho}$	ρ_{ij} 67% CI	90% CI	$\hat{\beta}$	β_{ij} 67% CI	90% CI	$\hat{\sigma}_{ij,y|x}$
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 6: continued
b. Selected other variables

Y	X	q	$i{:}j$	$\hat{\rho}$	ρ_{ij}		$\hat{\beta}$	β_{ij}		$\hat{\sigma}_{ij,y\|x}$
					67% CI	90% CI		67% CI	90% CI	
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
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
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
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

Notes: Results are shown for a subset of the variables listed in Tables 3-5. Results are based on Ω_{ij} (col. 4 lists i and j) and sample information in q cosine transforms (col. 3 shows q).

relationships involving the other variables. (Results for all the pairs of variables shown in Table 5 are available in the supplementary appendix.)

The first block of results in Table 6.a are for consumption and GDP. The first row repeats earlier results using the $q = 12$ cosine terms to learn about $\Omega_{i:j}$ with $i = 1$ and $j = 12$. The other rows are for other values of q, i, and j. The results suggest remarkable stability across the different values of q, i, and j. Figures 6.c and 6.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 6.a and 6.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 6. This same stability across q, i, and j is evident for the other pairs of variables in Table 6.a. Looking closely at Table 6.a, 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 $q = 12$.

The results summarized in Table 6.b show much of the same stability as Table 6.a, but there are some notable differences. For example, the point estimates suggest a somewhat larger Fisher correlation over longer periods (greater than 23 years) than over shorter periods (11 to 19 years), and the same holds for stock prices and dividends. In both cases however, the confidence intervals remain wide. And, the puzzling negative correlation between the unemployment rate and TFP appears to be stronger over the longer-long run than over the medium-long run.

7   Concluding remarks

This paper has focused on inference about long-run covariability of two time series. Just as with previous frameworks, such as cointegration analysis, 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

and population R^2s.[12] 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. And, confidence sets for multiple regression parameters satisfy natural invariance and equivariance constraints, which reduces the number of effective parameters.

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 $O(n^2 q^2)$ 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 is not too far below the nominal level, so these Bayes sets wouldn't be too misleading even from a frequentist perspective.[13] While this will be difficult to exhaustively check, this pattern might well generalize also to larger values of n.

---

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

[13]In fact, a calculation analogous to those in Table 2 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*.

CAMPBELL, J. Y. (1987): "Does Savings Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," *Ecomometrica*, 55, 1249–1273.

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.

COCHRANE, J. (1994): "Permanent and Transitory Components of GNP and Stock Prices," *Quarterly Journal of Economics*, CIX, 241–266.

COMIN, D., AND M. GERTLER (2006): "Medium Term Business Cycles," *American Economic Review*, 96, 523–551.

DAI, Q., AND K. SINGLETON (2000): "Specification analysis of affine term structure models," *Journal of Finance*, 55, 1943–1978.

DAVIDSON, J. E., D. F. HENDRY, F. SRBA, AND S. YEO (1978): "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal*, 88, 661–692.

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.

DIEBOLD, F. X., AND C. LI (2006): "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130, 337–364.

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.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

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. (1981): "Some Properties of Time Series Data and their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.

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.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

KIEFER, N. M., T. J. VOGELSANG, AND H. BUNZEL (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

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

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.

LAZARUS, E., D. J. LEWIS, AND J. STOCK (2016): "HAR Inference: Kernel Choice, Size Distortions, and Power Loss," *Working Paper, Harvard University*.

LETTAU, M., AND S. LUDVIGSON (2013): *Shocks and Crashes*. MIT Press, Cambridge MA.

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.

MÜLLER, U. K. (2004): "A Theory of Robust Long-Run Variance Estimation," *Working paper, Princeton University*.

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

———— (2014): "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, 32, 311–322.

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. (1991): "Optimal Inference in Cointegrated Systems," *Ecomometrica*, 59, 283–306.

PHILLIPS, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.

PRATT, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 549–567.

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.

SARGAN, J. D. (1964): "Wages and Prices in the UK: A Study in Econometric Methodology," in *Econometric Analysis for National Planning*, ed. by P. E. Hart, G. Mills, and J. K. Whitaker, pp. 25–26, London. Butterworths.

SHILLER, R. J. (2000): *Irrational Exuberance*. Princeton University Press, Princeton.

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.

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

SUN, Y. (2013): "Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator," *The Econometrics Journal*, 16, 1–26.

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

by Ulrich K. Müller and Mark W. Watson

This appendix provides supplemental material. Section 1 discusses the form of the confidence sets; section 2 derives the necessary densities; section 3 discusses the numerically determined approximate least favorable distributions; the data are described in section 4, and section 5 includes an expanded version of the paper's Table 6.

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 $\theta \in \Theta$.[14] Then

$$H_0^\rho(x,y) = \left\{ r : \alpha/2 \leq \frac{\int \mathbf{1}[g^\rho(\theta) \leq r] f^s(x^s, y^s|\theta) dF(\theta)}{\int f^s(x^s, y^s|\theta) dF(\theta)} \leq 1 - \alpha/2 \right\} \qquad \text{(A.1)}$$

$$H^\rho(x,y) = \left\{ r : \int f^s(x^s, y^s|\theta) dW(\theta) \geq \int f^s(x^s, y^s|\theta) d\Lambda_r^\rho(\theta) \right\} \cup H_0^\rho(x,y)$$

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 $\Lambda_r^\rho(\{\theta : g^\rho(\theta) \neq r$ 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 $\theta \in \Theta$.

---

[14]Here 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 $Z^* = (X_0^*, X^{*\prime}, Y^{*\prime})'$. Then

$$H_0^\beta(x, y) = \left\{ b : \alpha/2 \le \frac{\int \int \mathbf{1}[\frac{u_{11}b - u_{12}}{u_{22}} \le w] f_0^\beta(z^*, w|\theta) dw dF(\theta)}{\int f_1^*(z^*|\theta) dF(\theta)} \le 1 - \alpha/2 \right\}$$

$$H^\beta(x, y) = \left\{ b : \int h^\beta(z^*|\theta) f_1^*(z^*|\theta) dW(\theta) \ge \int f_0^\beta\left(z^*, \frac{u_{11}b - u_{12}}{u_{22}}|\theta\right) d\Lambda^\beta(\theta) \right\} \cup H_0^\beta(x, y)$$

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 $(Z^{*\prime}, (U_{11}g^\beta(\theta) - U_{12})/U_{22})'$ 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 $P_\theta(g^\beta(\theta) \in H^\beta(X, Y)) \ge 1 - \alpha$ for all $\theta \in \Theta$.

H^σ:

$$H_0^\sigma(x, y) = \left\{ s : \alpha/2 \le \frac{\int \int \mathbf{1}[\frac{s}{u_{22}} \le w] f_0^\sigma(z^*, w|\theta) dw dF(\theta)}{\int f_1^*(z^*|\theta) dF(\theta)} \le 1 - \alpha/2 \right\}$$

$$H^\sigma(x, y) = \left\{ s : \int h^\sigma(z^*|\theta) f_1^*(z^*|\theta) dW(\theta) \ge \int f_0^\sigma\left(z^*, \frac{s}{u_{22}}|\theta\right) d\Lambda^\sigma(\theta) \right\} \cup H_0^\sigma(x, y)$$

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 $(Z^{*\prime}, g^\sigma(\theta)/|U_{22}|)'$ 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 $P_\theta(g^\sigma(\theta) \in H^\sigma(X, Y)) \ge 1 - \alpha$ 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.

2 Densities of Maximal Invariants and Related Results

2.1 Preliminaries

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

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

$$= a^{p_1-1}b^{p_2-1}\frac{1}{2\pi}\int_0^\infty\int_0^\infty s^{p_1}t^{p_2}\exp[-\tfrac{1}{2}(s^2-2rst+t^2)]dsdt$$

$$= a^{p_1-1}b^{p_2-1}\int_0^\infty t^{p_2}\phi(\sqrt{1-r^2}t)\int_0^\infty s^{p_1}\phi(s-rt)dsdt$$

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

$$\frac{1}{2\pi}\int_{-\infty}^\infty\int_{-\infty}^\infty |s|^{p_1}|t|^{p_2}\exp[-\tfrac{1}{2}\begin{pmatrix}s\\t\end{pmatrix}'\begin{pmatrix}a^2 & abr\\ abr & b^2\end{pmatrix}\begin{pmatrix}s\\t\end{pmatrix}]dsdt$$

$$= 2Q(r)+2Q(-r).$$

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

(a) With $B_p(m)=\int_{-\infty}^\infty\phi(s-m)s^pds$, 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 $I_l(h)=\frac{1}{\Phi(h)}\int_{-\infty}^h\phi(s)s^lds$,

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

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

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 $\phi'(s)=-s\phi(s)$

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

and the result follows.

(b) See Dhrymes (2005).

3

(c) Immediate after substituting $s^2 \to u$ 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 $\phi'(s) = -s\phi(s)$

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

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

Now by Lemma B.1 (a),

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

for some polynomial C_0 whose coefficients may be determined explicitly by the formula in Lemma B.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 B.1 (b). The remaining integral over dt is of the form

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

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

In the following, we simply write Σ for the covariance matrix of $\mathrm{vec}(X, Y)$, 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

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

4

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

$$(2\pi)^{-q}|\det \Sigma|^{-1/2} \exp[-\tfrac{1}{2} \begin{pmatrix} x^s s_x \\ y^s s_y \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} x^s s_x \\ y^s s_y \end{pmatrix}] s_x^{q-1} s_y^{q-1}$$

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

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

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

2.3   Derivation of f_1^*

With $X^\dagger = (1, X_0^*, X^{*\prime})'$, $Y^\dagger = (1, 0, Y^{*\prime})'$ 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 $\mathbb{R}^{2q} \to \mathbb{R}^{2q}$ 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

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

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

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

5

$$
= \begin{pmatrix} u_{11} \\ 0 \\ u_{12} \\ u_{22} \end{pmatrix}' \begin{pmatrix} z^\dagger & 0 \\ 0 & z^\dagger \end{pmatrix}' \begin{pmatrix} \Sigma_{xx}^- & \Sigma_{xy}^- \\ \Sigma_{yx}^- & \Sigma_{yy}^- \end{pmatrix} \begin{pmatrix} z^\dagger & 0 \\ 0 & z^\dagger \end{pmatrix} \begin{pmatrix} u_{11} \\ 0 \\ u_{12} \\ u_{22} \end{pmatrix}
$$

$$
= \begin{pmatrix} u_{11} \\ u_{22} \\ u_{12} \end{pmatrix}' \begin{pmatrix} x^{\dagger\prime}\Sigma_{xx}^- x^\dagger & \cdot & \cdot \\ y^{\dagger\prime}\Sigma_{yx}^- x^\dagger & y^{\dagger\prime}\Sigma_{yy}^- y^\dagger & \cdot \\ x^{\dagger\prime}\Sigma_{yx}^- x^\dagger & y^{\dagger\prime}\Sigma_{yy}^- x^\dagger & x^{\dagger\prime}\Sigma_{yy}^- x^\dagger \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{22} \\ u_{12} \end{pmatrix}
$$

$$
= \begin{pmatrix} \hat{u} \\ u_{12} \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} \hat{u} \\ u_{12} \end{pmatrix}
$$

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

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

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

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

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

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

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

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

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

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

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

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 $V_w = \begin{pmatrix} I_2 \\ \lambda_w' \end{pmatrix}' \begin{pmatrix} V & v \\ v' & v_0^2 \end{pmatrix} \begin{pmatrix} I_2 \\ \lambda_w' \end{pmatrix}$,

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

7

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 $(Z^*, \tilde{W}^\sigma)$. 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 $\tilde{f}_0^\sigma$.

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 $(Z^*, \tilde{W}^\sigma)$ can thus be written as

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

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

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

$$
= \sqrt{2\pi}v_0^{-1}\exp[-\tfrac{1}{2} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}' (V - vv'/v_0) \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}]
$$

so that

$$
\tilde{f}_0^\sigma(z^*, \tilde{w}^\sigma | \theta) = (2\pi)^{-q+1/2}(\det \Sigma)^{-1/2}|g^\beta(\theta)|^{q-1}|\tilde{w}^\sigma|^{-q}v_0^{-1}
$$

$$
\times \int |u_{11}|^{q-1}\exp[-\tfrac{1}{2} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}' \tilde{V} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}]du_{11}.
$$

8

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

$$\int |u_{11}|^{q-1} \exp[-\tfrac{1}{2} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}' \tilde{V} \begin{pmatrix} u_{11} \\ g^\sigma(\theta)/\tilde{w}^\sigma \end{pmatrix}] du_{11}$$

$$= \int |u_{11}|^{q-1} \exp[-\tfrac{1}{2}(\tilde{v}_{11}^2 u_{11}^2 + 2b_w \tilde{v}_{11} u_{11} + a_w^2)] du_{11}$$

$$= \tilde{v}_{11}^{-q} \int |\omega|^{q-1} \exp[-\tfrac{1}{2}(\omega^2 + 2b_w \omega + a_w^2)] d\omega$$

$$= \tilde{v}_{11}^{-q} \exp[-\tfrac{1}{2}(a_w^2 - b_w^2)] \int |\omega|^{q-1} \exp[-\tfrac{1}{2}(\omega + b_w)^2] d\omega.$$

For $q-1$ even, a closed-form expression for the integral follows from Lemma B.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 B.1 (a) and (b).

3 Determination of Approximate Least Favorable Distributions

3.1 Overview

The algorithm is a modified version of what is suggested in Elliott, Müller, and Watson (2015). Let the set $\Theta_c = \{\theta_1, \ldots, \theta_m\} \subset \Theta$ 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 1.[15] 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^*),

---

[15]Here 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 paramter $\epsilon > 0$) 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.

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 $\eta = (\eta_1, \ldots, \eta_8) \in [0,1]^8$, we set

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

where $\mathrm{chol}(\cdot)$ is the Choleski decomposition of a matrix, $O(\phi_\eta)$ is the 2×2 rotation matrix for the angle ϕ_η, and $S_{c,d} = \mathrm{diag}(\sqrt{q/\mathrm{tr}\,\Sigma_X(c_1, d_1)}, \sqrt{q/\mathrm{tr}\,\Sigma_X(c_2, d_2)})$, 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

10

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 $c_j \to 0$ and $c_j \to \infty$.

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, $g^\rho(\theta) = r$.

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

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

3.4   Importance Sampling

For H^ρ, the approach is exactly as described in Elliott, Müller, and Watson (2015), that is we employ the importance sampling approximation

$$P_\theta(g^\rho(\theta) \notin H^\rho_{\Lambda_c^*}(X, Y)) \approx N^{-1} \sum_{i=1}^{N} \frac{f^s(X^s_{(i)}, Y^s_{(i)}|\theta)}{f^s_p(X^s_{(i)}, Y^s_{(i)})} \mathbf{1}[g^\rho(\theta) \notin H^\rho_{\Lambda_c^*}(X^s_{(i)}, Y^s_{(i)})] \tag{A.3}$$

for some proposal density f^s_p, where $(X^s_{(i)}, Y^s_{(i)})$ are i.i.d. draws from f^s_p. For given r, this obviously induces a smooth approximating function on Θ_r, since for all $\theta \in \Theta_r$, $g(\theta) = r$,

11

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 $X^\dagger = (1, X_0^*, X^{*\prime})'$ and $Y^\dagger = (1, 0, Y^{*\prime})'$. Thus, given that $(X^\dagger, Y^\dagger)$ are functions of Z^*, we have

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

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,

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

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, the importance sampling weights $f_0(Z^*, W | \theta)/f_p(Z^*, W)$ have a large variance $\mathrm{Var}[f_0(Z^*, W | \theta)/f_p(Z^*, W)]$ with (Z^*, W) drawn from f_p, leading to 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 employ the following iterative scheme to construct a suitable f_p that is a mixture of 400 values of θ. For simplicity, we describe it only in the notation that is relevant for H^β and H^σ:

1. Select 100 initial points θ_j based on a low-discrepancy sequence for $\eta \in [0, 1]^8$ of subsection 3.2. For each θ_j, draw 200 i.i.d. draws $(Z_{(i)}^*, W_{(i)})$ from $f_0(z^*, w | \theta_j)$.

2. While the number θ_j-vectors is no larger than 400:

   (a) Set f_p to the equal probability mixture of the current set of θ_js.

   (b) Use a gradient search algorithm to find θ_0 that maximizes the empirical counterpart of the importance sampling weight variance based on the current f_p and

12

draws $(Z^*_{(i)}, W_{(i)})$. We use the larger of 2 BFGS searches starting from random starting values.

(c) Draw 200 i.i.d. draws $(Z^*_{(i)}, W_{(i)})$ from $f_0(z^*, w|\theta_0)$, and add θ_0 to the set of θ_js.

To ensure unbiased importance sampling estimates, we discard the 400×200 draws of $(Z^*_{(i)}, W_{(i)})$ that were generated in this determination of f_p, and redraw $N = 250,000$ i.i.d. draws from f_p for the actual computation of non-coverage probabilities via (A.3), (A.4) and (A.5).

3.5   Computation of Credible Sets and Integrals over F

We approximate integrals over F by a discrete sum over 1000 points θ^F_j, 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 $(X^s_{(i)}, Y^s_{(i)})$ or $Z^*_{(i)}$. Rather, it suffices to determine whether $r \in H^\rho_0(X^s_{(i)}, Y^s_{(i)})$ or $W_{(i)} \in H_0(X^\dagger_{(i)}, Y^\dagger_{(i)})$, respectively. Under the discrete approximation to F, it hence suffices to check whether or not

$$\frac{\sum_{j=1}^{1000} \mathbf{1}[g^\rho(\theta^F_j) \leq r] f^s(X^s_{(i)}, Y^s_{(i)}|\theta^F_j)}{\sum_{j=1}^{1000} f^s(X^s_{(i)}, Y^s_{(i)}|\theta^F_j)} \tag{A.6}$$

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

$$\frac{\int \sum_{j=1}^{1000} \mathbf{1}[w \leq W^j_{(i)}] f_0(Z^*_{(i)}, w|\theta^F_j)dw}{\sum_{j=1}^{1000} f^*_1(Z^*_{(i)}|\theta^F_j)} \tag{A.7}$$

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 $(X^s_{(i)}, Y^s_{(i)})$ and $(Z^*_{(i)}, W^j_{(i)})$ 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 $g(\theta) \notin H_0$.

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%, respectively. 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 10 coverage violating points in up to 100 BFGS searches before augmenting Θ_c and recomputing Λ_c^*, which is fairly time consuming, especially if Θ_c consists of many points. Once 100 BFGS searches with up to 30 iterations each did not yield a violating point, we switch to BFGS searches with up to 100 iterations. The algorithm stops once 500 consecutive BFGS searches with up to 100 iterations each do not yield a coverage violating point.

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 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 10-20 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.

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×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.

4 Data Used

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

5 Additional Empirical Results

An extended version of Table 6 is given below.

Extended Table 6: Long-run covariation measures for selected variables with $\Omega_{i:j}$
Inference based on q cosine transforms.

Y	X	q	$i{:}j$	$\rho_{i:j}$			$\beta_{i:j}$			$\hat{\sigma}_{i:j,y\|x}$
				$\hat{\rho}$	67% CI	90% CI	$\hat{\beta}$	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

Extended Table 6: continued

Y	X	q	i:j	$\rho_{i:j}$			$\beta_{i:j}$			$\hat{\sigma}_{i:j,y|x}$
				$\hat{\rho}$	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

Extended Table 6: continued

Y	X	q	$i{:}j$	$\hat{\rho}$	$\rho_{i:j}$ 67% CI	90% CI	$\hat{\beta}$	$\beta_{i:j}$ 67% CI	90% CI	$\hat{\sigma}_{i:j,y|x}$
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

Extended Table 6: continued

Y	X	q	i:j	$\hat{\rho}$	$\rho_{i:j}$ 67% CI	90% CI	$\hat{\beta}$	$\beta_{i:j}$ 67% CI	90% CI	$\hat{\sigma}_{i:j,y|x}$
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

Extended Table 6: continued

Y	X	q	i:j	$\hat{\rho}$	$\rho_{i:j}$ 67% CI	90% CI	$\hat{\beta}$	$\beta_{i:j}$ 67% CI	90% CI	$\hat{\sigma}_{i,j,y|x}$
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

Extended Table 6: continued

Y	X	q	i:j	$\hat{\rho}$	$\rho_{i:j}$ 67% CI	90% CI	$\hat{\beta}$	$\beta_{i:j}$ 67% CI	90% CI	$\hat{\sigma}_{i:j,y|x}$
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 shown for a subset of the variables listed in Tables 3-5. Results are based on $\Omega_{i:j}$ (col. 4 lists i and j) 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.

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.