# Measuring Uncertainty about Long-Run Predictions*

Ulrich K. Müller and Mark W. Watson
Princeton University
Department of Economics
Princeton, NJ, 08544
February 2013 (Revised August 2013)

**Abstract**

Long-run forecasts of economic variables play an important role in policy, planning, and portfolio decisions. We consider long-horizon forecasts of average growth of a scalar variable, assuming that first differences are second-order stationary. The main contribution is the construction of predictive sets with asymptotic coverage over a wide range of data generating processes, allowing for stochastically trending mean growth, slow mean reversion and other types of long-run dependencies. We illustrate the method by computing predictive sets for 10 to 75 year average growth rates of U.S. real per-capita GDP, consumption, productivity, price level, stock prices and population.

**JEL classification:** C22, C53, E17

**Keywords:** low frequency, spectral analysis, prediction interval, least favorable distribution

*We thank Graham Elliott, James Stock, and Jonathan Wright for useful comments and advice. Support was provided by the National Science Foundation through grants SES-0751056 and SES-1226464.

# 1 Introduction

This paper is concerned with quantifying the uncertainty in long-run predictions of economic variables. Long-run forecasts and the uncertainty surrounding them play an important role in policy, planning, and portfolio decisions. For example, in the United States, an ongoing task of the Congressional Budget Office (CBO) is to forecast productivity and real GDP growth over a 75-year horizon to help gauge the solvency of the Social Security Trustfund. Uncertainty surrounding these forecasts is then translated into the probability of trust fund insolvency.[1] Inflation "Caps" and "Floors" are option-like derivatives with payoffs tied to the average value of price inflation over the next decade; their risk-neutral prices are determined by the probability that the long-run average of future values of inflation falls above or below a pre-specified threshold.[2] And, there is a large literature in finance discussing optimal portfolio allocations for long-run investors and how these portfolios depend on uncertainty in long-run returns.[3]

To be specific, let $x_t$ denote a time series, and suppose that data on $x_t$ exists for time periods $t = 1, ..., T$. Let $\overline{x}_{T+1:T+h} = h^{-1} \sum_{t=1}^{h} x_{T+t}$ denote the average value of the series between time periods $T + 1$ through $T + h$. We are interested in the date $T$ uncertainty about the value of $\overline{x}_{T+1:T+h}$, as characterized by prediction sets that contain $\overline{x}_{T+1:T+h}$ with a pre-specified probability (such as 90%). We structure the problem so that the coverage probability can be calculated using asymptotic approximations based on the central limit theorem. In particular we suppose that both $T$ and $h$ are large, and construct the prediction sets as a function of a relatively small number of weighted averages of the sample values of $x_t$. We apply a central limit theorem to the variable of interest ($\overline{x}_{T+1:T+h}$) and the predictors, and study an asymptotic version of the prediction problem based on the multivariate normal distribution. Were all the parameters of this normal distribution known (or consistently estimable), the prediction problem would be a straightforward application of optimal prediction in the multivariate normal model.

The problem is complicated by unknown parameters that characterize the stochastic process $x_t$ and hence also the normal distribution in the large-sample problem. We assume that the first differences $\Delta x_t = x_t - x_{t-1}$ are covariance stationary. Because we are interested in a long-run prediction ($\overline{x}_{T+1:T+h}$, for $h$ large) the parameters that matter are those that characterize the (pseudo-) spectrum of $x_t$ near frequency zero. Because of the paucity of

[1] See Congressional Budget Office (2005).

[2] See Kitsul and Wright (2012).

[3] See, for example, Campbell and Viceira (1999), Pastor and Stambaugh (2012), and Siegel (2007).

information in the sample about these low-frequency parameters, uncertainty about their values is an important component of the uncertainty about $\overline{x}_{T+1:T+h}$. Much of the paper is devoted to the problem incorporating uncertainty about these parameter values into the construction of prediction sets for $\overline{x}_{T+1:T+h}$. We begin by constructing a flexible parametric framework to characterize the long-run properties of the $x_t$ process and then translate uncertainty about the value of the parameters value into uncertainty about $\overline{x}_{T+1:T+h}$.

The low-frequency properties of $x_t$ are parameterized by $\{\mu, \sigma, \theta\}$ where $\mu$ describes the long-run level of the process, $\sigma$ the long-run scale, and $\theta$ the shape of the spectrum around frequency zero. We show how uncertainty about $\mu$ and $\sigma$ can be accounted for by considering predictions sets that are scale and location invariant. To motivate our parameterization of the local-to-zero frequency spectral shape, we consider three well-known models: the fractionally integrated model, parameterized by a scalar $d$ (where $d = 0$ denotes the $I(0)$ model and $d = 1$ denotes the $I(1)$ model); the local-level model, which is the sum of independent $I(0)$ and $I(1)$ processes and is parameterized by a scalar $b$, which measures the relative importance of the $I(0)$ process; and the diffusion (or local-to-unity AR) model parameterized by a scalar $c$, which measures the degree of mean reversion in the process.[4] We construct a three parameter model, $\theta = (b, c, d)'$, which nests these models and provides additional flexibility to characterize the low-frequency shape of the spectrum.

We use both Bayes and frequentist methods to construct prediction sets that reflect uncertainty about $\theta$. The Bayes procedure is conceptually straightforward: given a prior for $\theta$ and the Gaussianity of the limiting problem, the predictive density for $\overline{x}_{T+1:T+h}$ follows from Bayes rule, so that prediction sets are readily computed. The frequentist procedure instead constructs a prediction set that, by definition, controls coverage uniformly over all values

[4]The fractional $I(d)$ model generalizes the standard $I(d)$ model with integer $d$ to allow for more general persistence patterns. It has been used in economics to describe macroeconomic data on output and prices, interest rates, exchange rates, asset returns, and volatility; see Baillie (1996) and Robinson (2003) for surveys and introductions, and see Diebold and Inoue (2001) and Perron and Qu (2010) for the connection between the $I(d)$ and regime switching models. Harvey's (1989) local-level model (sometimes reparameterized as an IMA(1,1) model) and its generalizations have been used to model inflation (e.g., Nelson and Schwert (1977), Stock and Watson (2007), and Cogley, Primiceri, and Sargent (2009)), output (e.g., Harvey (1985), Watson (1986), Clark (1987)), output growth (e.g., Stock and Watson (1998)), the unemployment rate (e.g., Staiger and Stock (1997) and Gordon (1998)), asset returns and consumption growth rates (typically with the persistent component following an AR(1) process with AR-coefficient close to one as, for example, in Summers (1986), Bansal and Yaron (2004), and Pastor and Stambaugh (2012)). The local-to-unity model (Chan and Wei (1987) and Phillips (1987)) is widely used to model highly persistent/near-unit root processes including interest rates, (detrended) output, inflation, and exchange rates (e.g., Stock (1991)).

$\theta$. We show how frequentist prediction sets with small expected volume can be constructed using a "least favorable distribution" for $\theta$. By construction these frequentist prediction sets have attractive properties over repeated samples from the same data generating process. To also guarantee a sensible description of uncertainty conditional on any given sample, we modify these frequentist sets using insights from Müller and Norets (2012).

In economics, arguably the most well-known predictive densities and corresponding prediction sets are the "Rivers of Blood" shown in the Bank of England's *Inflation Report*. These are judgmental prediction sets for inflation that are computed over a four year horizon by the members of the Bank's Monetary Policy Committee. In contrast, we are interested in prediction sets over a much longer horizon (say 75 years) computed from probability models. Most of the existing literature on long-horizon forecasting stresses the difficulty of constructing good long-term forecasts under uncertainty about the long-run properties of the process. Granger and Jeon (2007) provide a mostly verbal account. Elliott (2006) compares alternative approaches to point forecasts and compares their mean squared errors. Kemp (1999), Phillips (1998) and Stock (1996, 1997) show that standard formulas for forecast uncertainty break down in the long-horizon local-to-unity model, but they do not provide constructive alternatives. In the related problem of estimating long-run impulse responses Pesavento and Rossi (2006) construct confidence sets that account for uncertainty about the local-to-unity parameter. Chapter 8.7 in Beran (1994) discusses forecasting of fractionally integrated series, and Doornik and Ooms (2004) use an ARFIMA model to generate long-run uncertainty bands for future inflation, but without accounting for parameter estimation uncertainty. Pastor and Stambaugh (2012) compute predictive variances of long-run forecasts of stock returns that account for parameter uncertainty in a Bayesian framework. For a recent discussion of long-run population forecasts and estimates of uncertainty see, for instance, Lee (2011). In this context, Raftery, Li, Sevcíkova, Gerland, and Heilig (2012) employ a Bayesian approach to describing uncertainty about future fertility rates.

To make better sense of the remainder of the paper, it is useful to consider the following simple "baseline" approach to constructing prediction sets: assume a pure fractional Gaussian process with parameter $-1/2 < d < 3/2$ for $x_t$, that is, $x_t = \mu + u_t$ with

$$\begin{aligned} &(1-L)^d u_t = \varepsilon_t \text{ for } -1/2 < d < 1/2 \\ &(1-L)^{d-1}\Delta u_t = \varepsilon_t \text{ for } 1/2 < d < 3/2, \text{ and } u_0 = 0 \\ &\varepsilon_t \sim iid\mathcal{N}(0,\sigma^2) \end{aligned} \tag{1}$$

and, after specifying some prior on the three parameters $(\mu, \sigma, d)$, construct a predictive set as the $1-\alpha$ highest posterior predictive density set for $\overline{x}_{T+1:T+h}$. This simple approach suffers

from five drawbacks, which we address as follows. First, model (1) imposes an extremely tight restriction on the shape of the (pseudo) spectrum of $x_t$, with both low- and high-frequency properties jointly governed by the single parameter $d$. In Section 2, we discuss how to extract the relevant low-frequency information about $x_t$ using a small number of trigonometrically weighted averages. By making our predictions sets depend on $x_t$ only through these low-frequency transforms, we avoid having to model the higher frequency properties of $x_t$. Second, one might question the choice of prior for the location and scale parameters $\mu$ and $\sigma$. This will be addressed by imposing appropriate invariance restrictions on the predictive sets, also discussed in Section 2. Third, the Gaussianity of the driving disturbances $\varepsilon_t$ in (1) is a strong assumption. We address this issue by deriving a joint central limit theorem for the low-frequency transforms of the observed data and $\overline{x}_{T+1:T+h}$ in Section 3. A fourth concern about model (1) is the restrictive nature of its single parameter description of low-frequency properties. This is addressed by the flexible three-parameter *bcd*-model, also developed in Section 3. A fifth and crucial concern about the baseline prediction sets concerns their Bayes nature: the credible sets for $\overline{x}_{T+1:T+h}$ have coverage $1-\alpha$ by construction only if the parameters are drawn from the prior distribution. This issue is taken up in Section 4, where we discuss frequentist prediction sets that achieve coverage uniformly over the relevant parameter space. Section 5 discusses implementation details, and quantifies the loss in prediction accuracy associated with using a small number of low-frequency weighted averages as predictors. Finally, in Section 6 we apply the methods to construct long-run prediction sets for several U.S. macroeconomic time series including real GDP, consumption, productivity, population growth rates, price inflation and stock market returns. We conclude in Section 7 with a short discussion of the small sample effect of heteroskedasticity on the coverage of the suggested prediction sets, as well as the challenges in extending the methods to multiple time series.

Because the analysis involves several steps, it is easy to lose sight of how the various steps lead to the prediction sets shown in Section 6. The appendix summarizes the key steps for computing long-run prediction sets. The summary serves as a guide for those who want to compute prediction sets and also serves to highlight the key formulae in the paper.

# 2 The Prediction Problem

Let $x_t$ be the growth rate of the economic variable of interest, which is observed for $t = 1, \ldots, T$. The objective is to construct prediction sets, denoted by $A$, of the average value

of $x_t$ from periods $T+1$ to $T+h$,

$$\overline{x}_{T+1:T+h} = h^{-1}\sum_{t=1}^{h} x_{T+t} \tag{2}$$

with the property that $P(\overline{x}_{T+1:T+h} \in A) = 1-\alpha$, where $\alpha$ is a pre-specified constant. (Predicting the average growth rate alternatively corresponds to predicting the future level of the economic variable at time $T+h$.)

*Cosine transformations of the sample data.* It will turn out to be convenient to transform the sample data $\{x_t\}_{t=1}^T$ into the weighted averages $(\overline{x}_{1:T}, X_T)$, with $X_T = (X_T(1),\ldots,X_T(T-1))'$, and where $X_T(j)$ is the $j$th cosine transformation

$$X_T(j) = \int_0^1 \Psi_j(s) x_{\lfloor sT\rfloor+1} ds = \iota_{jT} T^{-1}\sum_{t=1}^{T} \Psi_j\left(\frac{t-1/2}{T}\right) x_t \tag{3}$$

with $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$ and $\iota_{jT} = (2T/j\pi)\sin(j\pi/2T) \to 1$. We make two remarks about this transformation. First, because the $\Psi_j$ weights add to zero, $X_T(j)$ is invariant to location shifts of the sample. Second, the transformation isolates variation in the sample data corresponding to different frequencies: $\overline{x}_{1:T}$ captures 0-frequency variation and $X_T(j)$ captures variation at frequency $j\pi/T$.

*Truncating the information set.* We will construct prediction sets based on a truncated information set that includes $\overline{x}_{1:T} = T^{-1}\sum_{t=1}^T x_t$ and the first $q$ elements of $X_T$ (denoted by $X_{T,1:q}$) and where $q$ is much smaller than $T-1$, so that $A = A(\overline{x}_{1:T}, X_{T,1:q})$. We do this for two reasons. The first is tractability: with a focus on this truncated information set, the analysis involves a small number of variables (the $(q+2)$ variables $(\overline{x}_{T+1:T+h}, \overline{x}_{1:T}, X_{T,1:q})$), and because each of these variables is a weighted average of $\{x_t\}_{t=1}^{T+h}$, a central limit derived in the next section allows us to study a limiting Gaussian version of the prediction problem that is much simpler than the original finite-sample problem. The second motivation for truncating the information set is robustness: we use the low-frequency information in the sample data ($\overline{x}_{1:T}$ and the first $q$ elements of $X_T$) to inform us about a low-frequency, long-run average of future data, but we do not use high frequency sample information (the last $T-1-q$ elements of $X_T$). While high frequency information is informative about low-frequency characteristics for some stochastic processes (for example, tightly parameterized ARMA processes), this is generally not the case, and high-frequency sample variation may lead to faulty low-frequency inference. Müller and Watson (2008, 2012) discuss this issue in detail. In Section 5 below we present numerical calculations that quantify the efficiency-robustness trade-off in the long-run prediction problem considered here.

*Invariance.* In our applications it is natural to restrict attention to prediction sets that are invariant to location and scale, so for example, the results will not depend on whether the data are expressed as growth rates in percentage points at an annual rate or as percent per quarter. Thus, we restrict attention to prediction sets with the property that if $y \in A(\overline{x}_{1:T}, X_{T,1:q})$ then $m + by \in A(m + b\overline{x}_{1:T}, bX_{T,1:q})$ for any constants $m$ and $b \neq 0$ (where the transformation of $X_{T,1:q}$ does not depend on $m$ because $X_{T,1:q}$ is location invariant). Invariance allows us to restrict attention to prediction sets that depend on functions of the sample data that are scale and location invariant; in particular we can limit attention to constructing prediction sets for $Y_T^s$ given $X_{T,1:q}^s$, where $Y_T^s = Y_T / \sqrt{X_{T,1:q}' X_{T,1:q}}$ with

$$Y_T = \overline{x}_{T+1:T+h} - \overline{x}_{1:T} \tag{4}$$

and $X_{T,1:q}^s = X_{T,1:q} / \sqrt{X_{T,1:q}' X_{T,1:q}}$.[5]

# 3 Large-Sample Approximations and a Low-Frequency Parameterization

The last section laid out the finite-sample prediction problem, and this section develops a large-sample approximation to the problem. We do this in three steps. First, we discuss conditions on the stochastic process $x_t$ that yield a limiting normal distribution for appropriately scaled versions of $(X_{T,1:q}, Y_T)$, and we show how the covariance matrix of this limiting normal distribution depends on the low-frequency characteristics of the process $x_t$, specifically the shape of its spectrum near frequency zero. The second step is to parameterize this covariance matrix in terms of the shape of this "local-to-zero spectrum". We do so in a way that three important benchmark models, the $I(d)$, local-level and local-to-unity models emerge as special cases. In the final step we present the implied asymptotic distribution of the maximal invariants $Y_T^s$ and $X_T^s$ that define the prediction problem under location and scale invariance.

[5] Setting $m = -\overline{x}_{1:T} / \sqrt{X_{T,1:q}' X_{T,1:q}}$ and $b = 1/\sqrt{X_{T,1:q}' X_{T,1:q}}$ implies that for any invariant set $A$, $y \in A(\overline{x}_{1:T}, X_{T,1:q})$ if and only if $(y - \overline{x}_{1:T}) / \sqrt{X_{T,1:q}' X_{T,1:q}} \in A(0, X_{T,1:q} / \sqrt{X_{T,1:q}' X_{T,1:q}})$, and thus also $\overline{x}_{T+1:T+h} \in A(\overline{x}_{1:T}, X_{T,1:q})$ if and only if $Y_T^s \in A(0, X_{T,1:q}^s)$.

## 3.1 Asymptotic Behavior of $(X_{T,1:q}, Y_T)$

To derive the asymptotic behavior for $(X_{T,1:q}, Y_T)$, note that each element can be written as a weighted average of the elements of $\{x_t\}_{t=1}^{T+h}$. Thus, let $g : [0, 1+r] \mapsto \mathbb{R}$ denote a generic weighting function, where $r = \lim_{T\to\infty}(h/T) > 0$, and consider the weighted average of $\{x_t\}_{t=1}^{T+h}$

$$\eta_T = T^{1-\alpha} \int_0^{1+r} g(s) x_{\lfloor sT \rfloor + 1} ds$$

where $\alpha$ is a suitably chosen constant.[6] In our context, the elements of $X_{T,1:q}$ are cosine transformations of the in-sample values of $x_t$ (cf (3)), so that $g(s) = \sqrt{2}\cos(j\pi s)$ for $0 \leq s \leq 1$ and $g(s) = 0$ for $s > 1$; $Y_T$ defined in (4) is the difference between the out-of-sample and in-sample average values of $x_t$, so that $g(s) = -1$ for $0 \leq s \leq 1$ and $g(s) = r^{-1}$ for $1 < s \leq 1 + r$. These weights sum to zero, so that the (unconditional) expectation of $x_t$ plays no role in the study of $\eta_T$.

In the appendix we provide a central limit theorem for $\eta_T$ under a set of primitive conditions about the stochastic process describing $x_t$ and these weighting functions. We will not list the technical conditions in the text, but rather give a brief overview of the key conditions before stating the limiting result and discussing the form of the limiting covariance matrix. In particular, the analysis is carried out under the assumption that $\Delta x_t$ has the moving average representation $\Delta x_t = c(L)\varepsilon_t$, where $\varepsilon_t$ is a possibly conditionally heteroskedastic martingale difference sequence with more than 2 unconditional moments. (The restriction that $E\Delta x_t = 0$ rules out a linear time trend in $x_t$, but recall that $x_t$ denotes a growth rate, so this is not overly restrictive.) The moving average coefficients in $c(L)$ are square summable, so that $\Delta x_t$ has a spectrum, denoted by $F(\lambda)$.[7] Let $R(\lambda) = F(\lambda)/|1 - e^{-i\lambda}|^2$ denote the spectrum of $x_t$ if it exists and the pseudo-spectrum otherwise.

Under these and additional technical assumptions, Theorem 1 in the appendix shows

[6] For the $I(0)$ model $\alpha = 1/2$, for the $I(1)$ model $\alpha = 3/2$, and so forth. Because of scale invariance, the limit distribution of $(X^s_{T,1:q}, Y^s_T)$ derived in Section 3.3 does not depend on $\alpha$.

[7] In the appendix we assume that $\{\Delta x_t\}$ is generated by a triangular array, but we ignore the dependence of $x_t$, $c(L)$, and $F(\lambda)$ on $T$ here for notational convenience.

that $\eta_T$ has a limiting normal distribution,[8] and as an implication

$$T^{1-\alpha}\begin{bmatrix} X_{T,1:q} \\ Y_T \end{bmatrix} \Rightarrow \begin{bmatrix} X_{1:q} \\ Y \end{bmatrix} \sim \mathcal{N}(0,\Sigma) \tag{5}$$

with $X_{1:q} = (X_1, \ldots, X_q)'$. The limiting covariance matrix $\Sigma$ in (5) depends on the (pseudo-) spectrum $R$ of $x_t$, and the weight functions $g$ associated with the elements of $(X_{T,1:q}, Y_T)$.

## 3.2 The Local-to-Zero Spectrum

The appendix derives $\Sigma$ and shows that it depends exclusively on the low-frequency properties of $x_t$. To see why, consider a special case of our analysis in which $x_t$ is stationary with spectrum $R$. The $jk$'th element of $\Sigma$ is the limiting covariance of the two weighted averages $\eta_{j,T}$ and $\eta_{k,T}$, where

$$\eta_{l,T} = T^{1-\alpha}\int_0^{1+r} g_l(s)x_{[sT]+1}ds = T^{-\alpha}\sum_{t=1}^{\lfloor(1+r)T\rfloor}\tilde{g}_{l,t}x_t$$

with $\tilde{g}_{l,t} = T\int_{(t-1)/T}^{t/T} g_l(s)ds$. Thus, recalling that the $j$th autocovariance of $x_t$ is given by $\int_{-\pi}^{\pi} R(\lambda)e^{-\mathrm{i}\lambda j}d\lambda$, where $\mathrm{i} = \sqrt{-1}$, we obtain

$$\begin{aligned}
E(\eta_{j,T}\eta_{k,T}) &= T^{-2\alpha}\sum_{s,t=1}^{\lfloor(1+r)T\rfloor}\left(\int_{-\pi}^{\pi} e^{-\mathrm{i}\lambda(t-s)}R(\lambda)d\lambda\right)\tilde{g}_{j,t}\tilde{g}_{k,s} \\
&= T^{-2\alpha}\int_{-\pi}^{\pi} R(\lambda)\left(\sum_{s=1}^{\lfloor(1+r)T\rfloor}\tilde{g}_{k,s}e^{\mathrm{i}\lambda s}\right)\left(\sum_{t=1}^{\lfloor(1+r)T\rfloor}\tilde{g}_{j,t}e^{-\mathrm{i}\lambda t}\right)d\lambda \\
&= T^{1-2\alpha}\int_{-T\pi}^{T\pi} R(\omega/T)\left(T^{-1}\sum_{s=1}^{\lfloor(1+r)T\rfloor}\tilde{g}_{k,s}e^{\mathrm{i}\omega(s/T)}\right)\left(T^{-1}\sum_{t=1}^{\lfloor(1+r)T\rfloor}\tilde{g}_{j,t}e^{-\mathrm{i}\omega(t/T)}\right)d\omega \\
&\to \int_{-\infty}^{\infty} S(\omega)\left(\int_0^{1+r} g_k(s)e^{\mathrm{i}\omega s}ds\right)\left(\int_0^{1+r} g_j(s)e^{-\mathrm{i}\omega s}ds\right)d\omega
\end{aligned} \tag{6}$$

where

$$S(\omega) = \lim_{T\to\infty} T^{1-2\alpha}R(\omega/T)$$

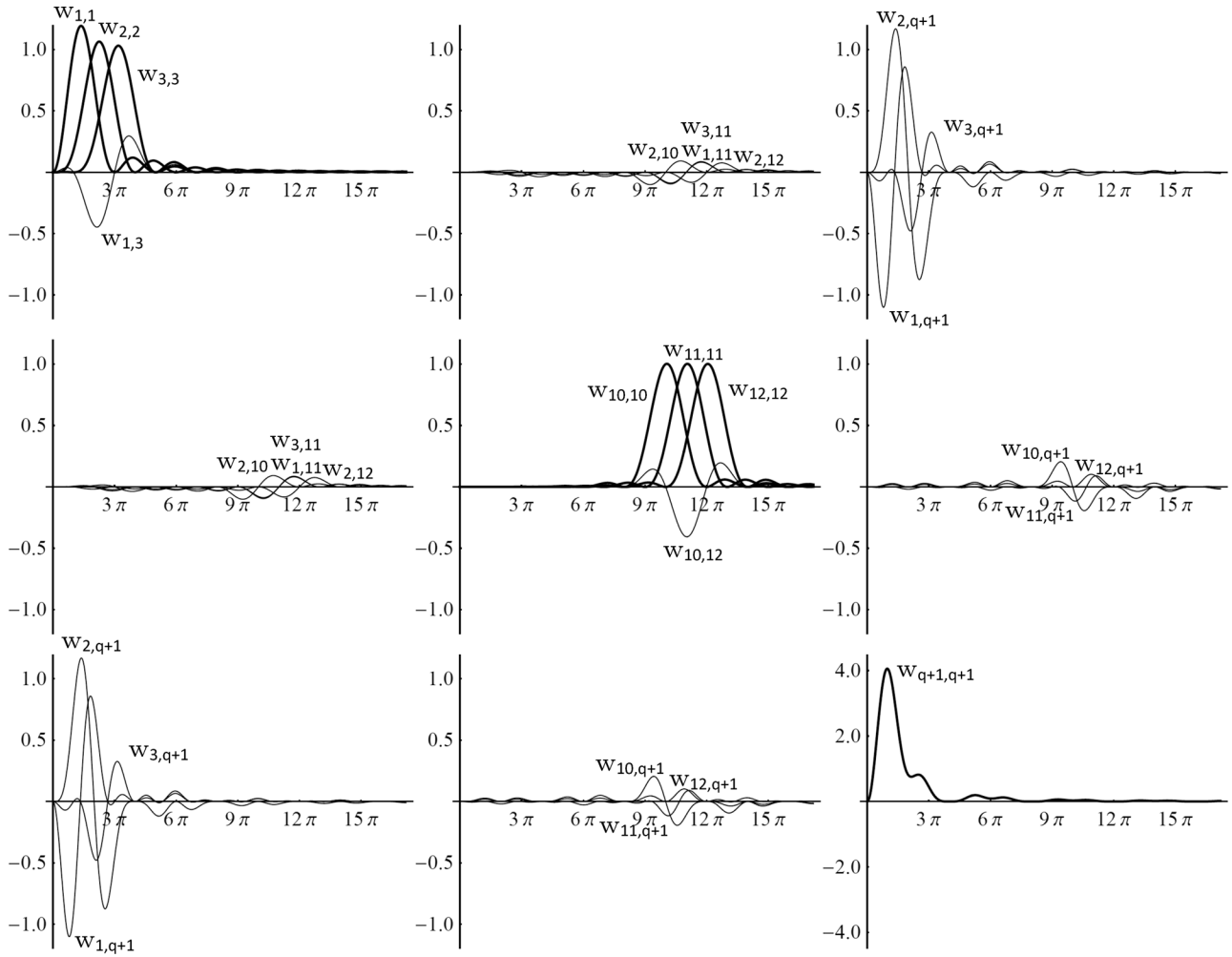
[8] As in any central limit theorem, the conditions underlying Theorem 1 imply that no single shock has a substantial impact on the overall variability of $\eta_T$. This assumption might be violated by rate but catastrophic events stressed in the work of Rietz (1988) and Barro (2006), for example. Note, however, that such events would need to substantially impact the *average* growth rate over a long horizon to invalidate a normal approximation.

is the 'local-to-zero spectrum' of $x_t$. The appendix shows that covariance stationarity of $x_t$ is not required for (6) to hold as long as $\Delta x_t$ is covariance stationary, in which case $S$ becomes the local-to-zero limit of the pseudo-spectrum $R$ of $x_t$.

The limit (6) shows that the elements of the covariance matrix of $(X_{1:q}, Y)'$ are simply weighted averages of the local-to-zero spectrum $S$. Note that since $S$ is an even function and $g_j$ and $g_k$ are real valued, (6) can be rewritten as $E(\eta_{j,T}\eta_{k,T}) \to 2\int_0^\infty S(\omega)w_{jk}(\omega)d\omega$, where $w_{jk}(\omega) = \mathrm{Re}[\left(\int_0^{1+r} g_j(s)e^{-\mathrm{i}\omega s}ds\right)\left(\int_0^{1+r} g_k(s)e^{\mathrm{i}\omega s}ds\right)]$. With $g_j = \sqrt{2}\cos(\pi j s)$, a calculation shows that $w_{jk}(\omega) = 0$ for $1 \le j, k \le q$ and $j+k$ odd, so that $E(X_jX_k) = 0$ for all odd $j+k$, independent of the local-to-zero spectrum $S$. Figure 1 plots $w_{jk}(\cdot)$ for some selected values of $j$ and $k$ and $r = 1/2$. The figure displays the weights $w_{j,k}$ corresponding to the covariance matrix of the vector $(X_1, X_2, X_3, X_{10}, X_{11}, X_{12}, Y)'$, organized into a symmetric matrix of 9 panels. The first panel plots the weights for the covariance matrix of $(X_1, X_2, X_3)'$, where the weights for the variances are shown in bold and the weights for the covariances are shown as thin curves. The second panel plots the weights associated with covariances between $(X_1, X_2, X_3)'$ and $(X_{10}, X_{11}, X_{12})'$, and so forth. A calculation shows that covariance weights $w_{jk}$, $j \neq k$, integrate to zero, which implies that for a flat local-to-zero spectrum $S$, $\Sigma$ is diagonal. As can be seen from the first and second panel on the diagonal of Figure 1, the variance of the predictor $X_j$ is mostly determined by the values of $S$ in the interval $\pi j \pm 2\pi$. Further, as long as $S$ is somewhat smooth, the correlation between $X_j$ and $X_k$ is very close to zero for $|j-k|$ large. The final panel in the figure shows the weight associated with the variance of $Y$, the variable being predicted. Evidently the unconditional variance of $Y$ is mostly determined by the shape of $S$ on the interval $\omega \in [0, 4\pi]$, and its correlation with $X_j$ is therefore small for $j$ large even for smooth but non-constant $S$ (a calculation shows that $\max_\omega |w_{q+1,j}(\omega)|$ decays at the rate $1/j$ as $j \to \infty$). The implication of these results is that the conditional variance of $Y$ given $X_{1:q}$ depends on the local-to-zero spectrum, with the shape of $S$ for, say, $\omega < 12\pi$, essentially determining its value, even for large $q$. In terms of the original time series, frequencies of $|\omega| < 12\pi$ correspond to cycles of periodicity $T/6$. For instance, with 60 years worth of data (of any sampling frequency), the shape of the spectrum for frequencies below 10 year cycles essentially determines the uncertainty of the forecast of mean growth over the next 30 years.

Since the local-to-zero spectrum $S$ determines $\Sigma$, the salient feature of the stochastic process $x_t$ for long-run forecasts is the shape of the local-to-zero spectrum (the unconditional mean of $x_t$, as well as the absolute scale of the local-to-zero spectrum of $x_t$ are irrelevant under the invariance requirement discussed in Section 2). To motivate our parameterization

Figure 1: Integral Weights on Local-to-Zero Spectrum for Selected Elements of Σ




Notes: The curves denote weights for variances (bold) and covariances (thin) of ($X_1$, $X_2$, $X_3$, $X_{10}$, $X_{11}$, $X_{12}$, $Y$) with $w_{j,q+1}$ corresponding to $E(X_jY)$, and $w_{q+1,q+1}$ corresponding to $E(Y^2)$. The nine panels show weights for the covariances of the three sets of variables ($X_1$, $X_2$, $X_3$), ($X_{10}$, $X_{11}$, $X_{12}$), and $Y$, respectively. All weights $w_{j,k}$ on covariance terms $\Sigma_{jk} = E(X_jX_k)$ with $j,k \leq q$ and $j+k$ odd are equal to zero.

of $S$, it is useful to consider three benchmark models that have been proven useful to model low-frequency phenomena in other contexts: the fractional $I(d)$ model, the local-level model, and the local-to-unity model. The $I(d)$ model has a (pseudo-) spectrum proportional to $|\lambda|^{-2d}$ for $\lambda$ close to zero, so that $S(\omega) \propto |\omega|^{-2d}$, $-1/2 < d < 3/2$. In particular $S$ is constant for the $I(0)$ model and is proportional to $\omega^{-2}$ for the $I(1)$ model. The local-level model expresses $x_t$ as the sum of an $I(0)$ process and an $I(1)$ process, say $x_t = e_{1t} + (bT)^{-1}\sum_{s=1}^{t} e_{2s}$, where $\{e_{1t}\}$ and $\{e_{2t}\}$ are mutually uncorrelated $I(0)$ processes with the same long-run variance. The second component of relative magnitude $1/b$ is usefully thought of as a stochastically varying 'local mean' of the growth rate $x_t$. In this model, $S(\omega) \propto b^2 + \omega^{-2}$. Finally, in the local-to-unity model $x_t = (1 - c/T)x_{t-1} + e_t$, where $e_t$ is an $I(0)$ process, and a straightforward calculation shows that $S(\omega) \propto 1/(\omega^2 + c^2)$. (Note that $(b, c) \to (\infty, \infty)$ and $(b, c) \to (0, 0)$ recover the $I(0)$ and $I(1)$ model, respectively). These three models are nested in the parameterization

$$S(\omega) \propto \left(\frac{1}{\omega^2 + c^2}\right)^d + b^2 \tag{7}$$

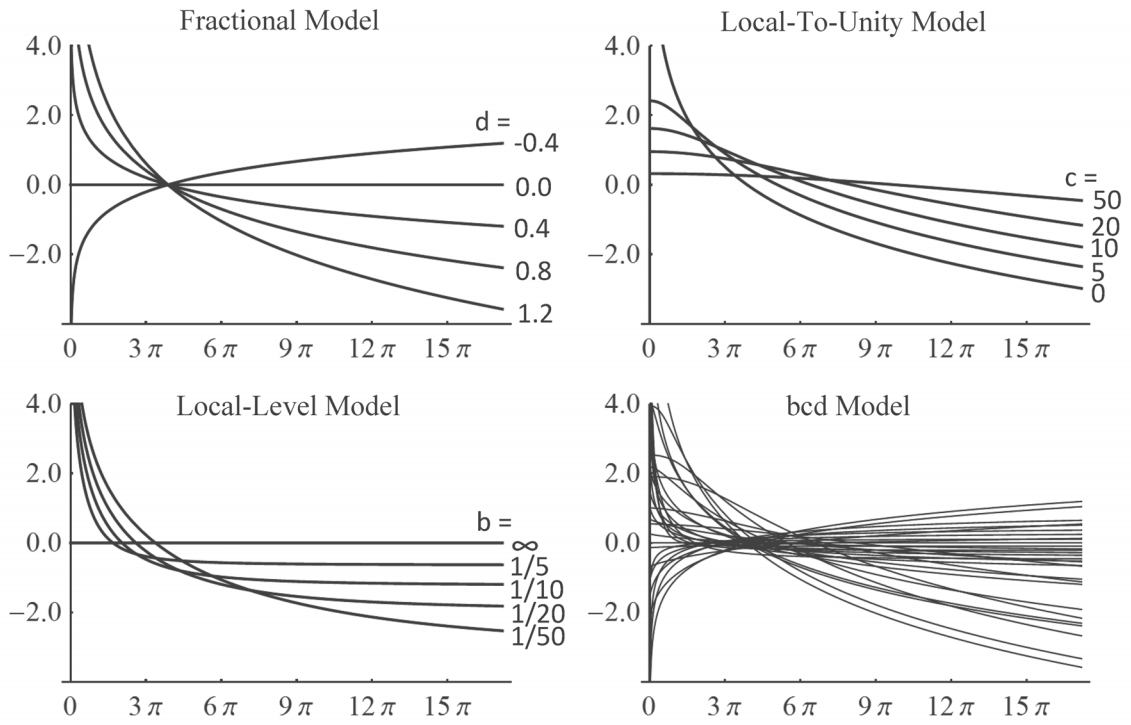
where $b = c = 0$ for the $I(d)$ model, $d = 1$, and $c = 0$ for the local-level model, and $d = 1$, and $b = 0$ for the local-to-unity model.[9]

In the construction of our prediction sets, we assume $S$ to follow this "*bcd*-model", with $-1/2 < d < 3/2$ and $b$ and $c$ unrestricted. This parameterization allows us to capture a wide range of monotone shapes for the low frequency (pseudo-) spectrum of $x_t$, including, but not limited to, the three benchmark models discussed above. Figure 2 shows the logarithm of the local-to-zero spectra of the three benchmark models for selected values of $d$, $b$ and $c$, and for several values of $b, c, d$ in the *bcd*-model. Each of the benchmark models can capture $I(1)$ and $I(0)$ shapes, and the $I(d)$ model can capture more than than $I(1)$ and less than $I(0)$ persistence. The last panel in the figure shows that the *bcd*-model captures a wide range of low-frequency shapes.[10]

[9] This is recognized as the local-to-zero spectrum of the process $x_t = e_{1t} + (bT^d)^{-1}z_t$, where $(1-\rho_T L)^d z_t = e_{2t}$ with $\rho_T = 1 - c/T$ and $\{e_{1t}\}$ and $\{e_{2t}\}$ are mutually uncorrelated $I(0)$ processes with the same long-run variance. This process is investigated further in Müller and Watson (2013).

[10] Note that the relative large values of $\log(S(\omega))$ for $|\omega|$ small in many parameter configurations imply an extreme loading on very low frequencies of $S$. The corresponding $\Sigma$ can nevertheless be computed by the integral of the product of the implied $S$ and the weight functions $w_{jk}$ of Figure 1, $\Sigma_{jk} = 2\int_0^\infty S(\omega)w_{jk}(\omega)d\omega$ (as we show in the appendix, it suffices that $\int \omega^2 S(\omega)d\omega < \infty$, which holds as long as $-1/2 < d < 3/2$ in the *abd*-model).

Figure 2: Logarithm of the Local-to-Zero Spectrum for Selected Models




Note: The scale of the local-to-zero spectrum S is normalized such that log(S(π)) + log(S(15π)) = 0.

### 3.3 The Limiting Density of $(X^s_{T,1:q}, Y^s_T)$

The limiting density of the invariants $X^s_{T,1:q} = X_{T,1:q}/\sqrt{X'_{T,1:q}X_{T,1:q}}$ and $Y^s_T = Y_T/\sqrt{X'_{T,1:q}X_{T,1:q}}$ follows directly from (5) and the continuous mapping theorem,

$$\begin{bmatrix} X^s_{T,1:q} \\ Y^s_T \end{bmatrix} \Rightarrow \begin{bmatrix} X^s \\ Y^s \end{bmatrix}$$

where $X^s = X_{1:q}/\sqrt{X'_{1:q}X_{1:q}}$ and $Y^s = Y/\sqrt{X'_{1:q}X_{1:q}}$ (we omit the dependence of $X^s$ on $q$ to ease notation). Let $\nu_q$ be the surface measure of a $q$ dimensional unit sphere, and $\mu_1$ Lebesgue measure on $\mathbb{R}$. Setting $W^s = (X^{s\prime}, Y^s)'$, the densities of $X^s$ and $W^s$ relative to $\nu_q$ and $\nu_q \times \mu_1$, respectively, are given by

$$f_{X^s}(x^s) = \tfrac{1}{2}\Gamma(\tfrac{q}{2})\pi^{-q/2}|\Sigma_X|^{-1/2}(x^{s\prime}\Sigma_X^{-1}x^s)^{-q/2} \quad (8)$$

$$f_{W^s}(w^s) = \tfrac{1}{2}\Gamma(\tfrac{q+1}{2})\pi^{-(q+1)/2}|\Sigma|^{-1/2}(w^{s\prime}\Sigma^{-1}w^s)^{-(q+1)/2} \quad (9)$$

where $\Sigma_X = E(X_{1:q}X'_{1:q})$. The result for $X^s$ is well-known (see Kariya (1980) or King (1980)), and the density of $W^s$ is derived in the appendix.

An important special case of this result is when $x_t$ is $I(0)$. In this case $S(\omega)$ is a constant and

$$\Sigma_{I(0)} \propto \begin{bmatrix} I_q & 0 \\ 0 & 1 + r^{-1} \end{bmatrix},$$

so that $f_{W^s}(w^s) \propto (1 + (y^s)^2/(1 + r^{-1}))^{-(q+1)/2}$. In the $I(0)$ model, $X^s$ and $Y^s$ are thus independent, and $q^{1/2}(1 + r^{-1})^{-1/2}Y^s$ is distributed Student-t with $q$ degrees of freedom.

# 4 Bayes and Frequentist Prediction Sets

Section 2 set up the prediction problem for $\overline{x}_{T+1:T+h}$, and Section 3 derived the limiting version of the problem in terms of $(X^s, Y^s)$. In this section we discuss Bayes and frequentist approaches for constructing prediction sets for $Y^s$ as functions of $X^s$. We consider a standard formulation of the problem (Pratt (1961)), namely, construction of a set $A(X^s)$ that has smallest expected volume subject to a coverage constraint

$$\min_A E(\text{vol}(A(X^s))) \text{ subject to } P(Y^s \in A(X^s)) \geq 1 - \alpha$$

where $\text{vol}(A)$ denotes the volume of the set $A$. The Bayes and frequentist versions of the problems use different distributions to compute the expected value in the objective function

and coverage probability in the constraint and therefore yield different solutions to the problem.

To discuss these solutions and how they differ, it is useful to introduce some generic notation: let $V$ denote a random variable with probability density $f_V(\cdot|\theta)$ that depends on $\theta$, let $\Omega$ denote a probability distribution for $\theta$, and let

$$f_V^{\Omega}(v) = \int f_V(v|\theta) d\Omega(\theta)$$

so that $f_V^{\Omega}$ is the density of $V$ if $\theta$ is first drawn randomly from $\Omega$.

## 4.1 Bayes Sets

In the Bayes version of the problem, the predictive density of $Y^s$ given $X^s = x^s$ is used to compute both the value of the objective function, $E(\text{vol}(A(X^s)))$ and the coverage constraint, $P(Y^s \in A(X^s)) \geq 1 - \alpha$. The resulting predictive set is characterized by the "highest predictive density" set

$$A_{\Gamma}^{\text{Bayes}}(x^s) = \{y^s : \frac{f_{(Y^s,X^s)}^{\Gamma}(y^s, x^s)}{f_{X^s}^{\Gamma}(x^s)} > \text{cv}(x^s)\} \tag{10}$$

where $f_{(Y^s,X^s)}^{\Gamma}(\cdot, x^s)/f_{X^s}^{\Gamma}(x^s)$ is the Bayes predictive density of $Y^s$ given $X^s = x^s$ using the prior $\Gamma$, and $\text{cv}(x^s)$ is a critical value associated with the coverage constraint (which depends on $x^s$ because the constraint is evaluated using the predictive density evaluated at $X^s = x^s$).

An important special case is Bayes forecasting in the $I(0)$ model (that is, $\Gamma$ puts all weight on the $I(0)$ model). Recall from Section 3 that in the $I(0)$ model, $Y^s$ is independent of $X^s$, and $q^{1/2}(1 + r^{-1})^{-1/2}Y^s$ is distributed Student-t with $q$ degrees of freedom. Letting $t_{(1-\alpha/2)}^q$ denote the $1 - \alpha/2$ quantile of this distribution, the $(1 - \alpha)$ Bayes prediction set for $Y^s$ is therefore $\{y^s : |y^s| \leq t_{(1-\alpha/2)}^q q^{-1/2}(1 + r^{-1})^{1/2}\}$, which does not depend on $x^s$. The corresponding predictive set for $\overline{x}_{T+1:T+h}$ is $\overline{x}_{1:T} \pm t_{(1-a/2)}^q(1 + r^{-1})^{1/2}T^{-1/2}s_{LR}$, where $s_{LR}^2 = (T/q)X_{T,1:q}'X_{T,1:q}$ is an estimator of the long-run variance of $x_t$.

## 4.2 Frequentist Sets

While the Bayes set enforces the coverage constraint conditional on each $X^s$, it averages over possible values of $\theta$ using the prior $\Gamma$. In contrast, the frequentist set enforces the coverage constraint for all values of $\theta$, but averages over possible values of $X^s$. The problem of constructing optimal frequentist prediction sets is analogous to the problem of constructing

optimal frequentist confidence sets, which in turn is closely related to the problem of constructing most powerful tests (Pratt (1961)). It is therefore useful to review the problem of constructing most powerful tests, as we employ analogous methods to construct optimal prediction sets.

In the testing problem, $\theta$ corresponds to a vector of parameters that characterize the distribution of the observables under the null hypothesis $H_0 : \theta \in \Theta_0$ and alternative hypothesis $H_1 : \theta \in \Theta_1$. Two devices are often used to handle composite null and alternative hypotheses. Under the alternative, the power of the test depends on the value of $\theta$, and tests are usefully evaluated by their weighted average power (WAP), where the weights load on different values of $\theta$. Thus, let $\Gamma : \Theta_1 \mapsto \mathbb{R}$ denote the weight function used to compute $\text{WAP}_\Gamma$ where the dependence on $\Gamma$ is made explicit. (While we have used the same notation $\Gamma$ for this weight function as for the prior in the Bayes problem, these represent conceptually different distributions.) For a level $\alpha$ test, the test's rejection frequency must be less than $\alpha$ uniformly over all values of $\theta \in \Theta_0$. This uniform size constraint can sometimes be enforced by assuming that $\theta$ is drawn from a "least favorable distribution," say $\Lambda$, on $\Theta_0$. When this is the case, the Neyman-Pearson lemma implies that the optimal (that is, the highest $\text{WAP}_\Gamma$) test can be constructed from the likelihood ratio, where the likelihood under the null is computed using $\Lambda$ and the likelihood under the alternative is computed using $\Gamma$. Elliott, Müller, and Watson (2013) discuss numerical methods for computing an approximate least favorable distribution $\tilde{\Lambda}$, which yields a value of $\text{WAP}_\Gamma$ that is approximately equal to its highest achievable value.

The prediction set analogue for these testing results is as follows. The objective function changes from maximizing power to minimizing expected volume, and because the distribution of $(Y^s, X^s)$ depends on $\theta$, the expected volume depends on $\theta$. Let $\text{WAV}_\Gamma$ denote the weighted average expected volume using the weight function $\Gamma$. The level constraint in the testing problem changes to the coverage constraint in the prediction set problem, and this is enforced uniformly over all values of $\theta$. Like the testing problem, this can sometimes be achieved using a least favorable distribution for $\theta$, and we therefore apply numerical procedures like those proposed in Elliott, Müller, and Watson (2013) to construct an approximate least favorable distribution for $\theta$, $\tilde{\Lambda}$, that guarantees uniform coverage. This allows us to construct an approximate optimal frequentist prediction set as the solution to the problem of minimizing $\text{WAV}_\Gamma$ subject to $P_{\tilde{\Lambda}}(Y^s \in A(X^s)) \geq 1-\alpha$, where $P_{\tilde{\Lambda}}$ means that the probability is computed by first drawing $\theta$ from $\tilde{\Lambda}$, and then $(Y^s, X^s)$ conditional on that value of $\theta$. The resulting

optimal prediction set is

$$A_{\Gamma}^{\text{freq}}(x^s) = \{y^s : \frac{f_{(Y^s,X^s)}^{\tilde{\Lambda}}(y^s, x^s)}{f_{X^s}^{\Gamma}(x^s)} > \text{cv}\}. \tag{11}$$

(Additional details are provided in the appendix.)

As a practical matter, the frequentist prediction set differs from the Bayes set in two ways. First, the Bayes set uses the same distribution of $\theta$ (the prior $\Gamma$) to evaluate the marginal densities of $X^s$ and $(Y^s, X^s)$, while the frequentist set uses different distributions ($\Gamma$ is used for the density of $X^s$ because this term arises from the expected volume calculation, and $\tilde{\Lambda}$ is used for $(Y^s, X^s)$ because this term arises from the coverage constraint). Second, the Bayes set uses a critical value that depends on $x^s$ (which enforces the coverage constraint conditional on $X^s = x^s$), while the frequentist set uses a fixed critical value (which enforces the coverage constraint averaging over all values of $X^s$ under $P_{\tilde{\Lambda}}$).

In the $I(0)$ model, $\theta$ is fixed, so the prior and weighting functions $\Gamma$ and $\tilde{\Lambda}$ are absent. In addition, the Bayes critical value does not depend on $x^s$ (because $Y^s$ and $X^s$ are independent). Thus, in the $I(0)$ model, the Bayes and frequentist prediction sets are identical.

### 4.2.1 Bet-proof Frequentist Prediction Sets

On average, frequentist prediction sets (11) have attractive properties: they cover $Y^s$ with a prespecified probability for all possible values of $\theta$, $\inf_{\theta\in\Theta} P_\theta(Y^s \in A_{\Gamma}^{\text{freq}}(X^s)) \geq 1-\alpha$, and they are as short as possible in a well defined sense. Conditional on a specific draw of $X^s = x^s$, however, they do not necessarily provide a compelling description of uncertainty about $Y^s$. To see this, consider the special case where $\theta = \theta_0$ is known (so that $\Gamma$ and $\tilde{\Lambda}$ in (10) and (11) put all mass at $\theta = \theta_0$), but the conditional density of $Y^s$ given $X^s = x^s$ depends on $x^s$. In our context, this would arise, for instance, if the $x$ process is known to be $I(1)$. A reasonable description of level $1-\alpha$ uncertainty about $Y^s$ after observing $X^s = x^s$ then arguably requires that the set covers (at least) $1-\alpha$ of the mass of the conditional density of $Y^s$ given $X = x^s$. The Bayes set $A_{\Gamma}^{\text{Bayes}}$ in (10) does so by choosing the shortest such set. In contrast, since cv in (11) is fixed, $A_{\Gamma}^{\text{freq}}$ does not have this property in general: for some realizations $X^s = x^s$, $P_{\theta_0}(Y^s \in A_{\Gamma}^{\text{freq}}(x^s)|X^s = x^s) > 1-\alpha$, for others, $P_{\theta_0}(Y^s \in A_{\Gamma}^{\text{freq}}(x^s)|X^s = x^s) < 1-\alpha$, and only the overall average over $X^s$, $P_{\theta_0}(Y^s \in A_{\Gamma}^{\text{freq}}(X^s))$, equals $1-\alpha$ by construction. In fact, $A_{\Gamma}^{\text{freq}}(x^s)$ might even by empty for some $X^s = x^s$. While this makes sense in the context of the frequentist problem—an empty set has no volume leading to a small value of the objective function, and coverage can be enforced by averaging over other values of $X^s$—it does not yield a reasonable measure of uncertainty about $Y^s$ after observing $X^s = x^s$.

Müller and Norets (2012) take up this issue in the context of frequentist confidence intervals. They consider a game where an inspector may bet that the realized confidence interval does not contain the true value, with a payoff that corresponds to the odds implied by the confidence level. Müller and Norets (2012) call a confidence set "bet-proof" if it is impossible for the inspector to generate positive expected winnings uniformly over the parameter space. It is easy to see that confidence sets that are empty with positive probability for all parameter values cannot be bet-proof, as the inspector could bet against empty realizations, and always be right. More generally, bet-proofness rules out descriptions of uncertainty that do not make sense conditionally, such as in the known parameter example above: whenever $P_{\theta_0}(Y^s \in A_\Gamma^{\text{freq}}(x^s)|X^s = x^s) < 1-\alpha$, the inspector bets that $Y^s \notin A_\Gamma^{\text{freq}}(x^s)$, which yields positive expected winnings.

It turns out that any bet-proof confidence or prediction set is necessarily a superset of a Bayes set of the same level, relative to some prior. One way of constructing attractive bet-proof frequentist sets suggested by Müller and Norets (2012) is thus to exogenously fix some prior $\Gamma$, and to then derive a $\Gamma$-weighted average expected volume minimizing set, subject to inclusion of a Bayes set relative to $\Gamma$, and the coverage constraint. We apply this suggestion here, with the Bayes set chosen as the equal-tailed $1-\alpha$ Bayes predictive interval. This choice is numerically more convenient than the highest predictive density set (10), and it induces bet-proofness in the extended game where the inspector may bet that $Y^s$ will be below (or above) the reported interval at odds corresponding to $\alpha/2$, as discussed by Müller and Norets (2012).

The numerical determination of such a bet-proof $\Gamma$-weighted average volume minimizing set is not much harder than what is described in Section 4.2 above, as the solution (11) is "realization by realization", and the additional constraint of the inclusion of the Bayes set just restricts $A_\Gamma^{\text{freq}}$ in (11) to be a superset of the equal-tailed Bayes predictive set. The approximately $\text{WAV}_\Gamma$ minimizing bet-proof frequentist prediction set is thus of the form

$$A_\Gamma^{MN}(x^s) = [q_{\alpha/2}^\Gamma(x^s); q_{1-\alpha/2}^\Gamma(x^s)] \cup \{y^s : \frac{f_{(Y^s,X^s)}^{\tilde{\Lambda}}(y^s,x^s)}{f_{X^s}^\Gamma(x^s)} > \text{cv}\} \tag{12}$$

where $q_{\alpha/2}^\Gamma(x^s)$ and $q_{1-\alpha/2}^\Gamma(x^s)$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the Bayes predictive density $f_{(Y^s,X^s)}^\Gamma(\cdot,x^s)/f_{X^s}^\Gamma(x^s)$, and $\tilde{\Lambda}$ and cv are generally different from their values in the unconstrained solution (11). By construction, the $A_\Gamma^{MN}$ prediction set has a frequentist interpretation (that is, it provides coverage of at least $1-\alpha$ averaged over $X^s$ for all values of $\theta$), but as superset of a Bayes sets, it also has coverage of at least $1-\alpha$ conditional on $X^s = x^s$ when $\theta$ is randomly drawn from the prior $\Gamma$. In the special case where $\theta$ is known, $A_\Gamma^{MN}$

reduces to the equal-tailed predictive set $[q^{\Gamma}_{\alpha/2}(x^s); q^{\Gamma}_{1-\alpha/2}(x^s)]$, since unconditional coverage is implied by the conditional coverage of $[q^{\Gamma}_{\alpha/2}(x^s); q^{\Gamma}_{1-\alpha/2}(x^s)]$ for all $X^s = x^s$. Further, since the conditional distribution of $Y^s$ given $X^s = x^s$ and $\theta = \theta_0$ is symmetric around zero in our prediction problem, it is then also equal to the Bayes highest predictive density set, $A^{\text{Bayes}}_{\Gamma}$.

# 5 Empirical Implementation

In this section, we bring together the results of the previous sections and discuss our choices for the weighting function $\Gamma$ and the number of predictors $q$ in the empirical implementation.

In our context, the parameter space consists of the possible values of $\theta = (b, c, d)'$, the parameter that describes the local-to-zero spectrum $S$ introduced in Section 3 above. We specify $-0.4 \leq d \leq 1.4$, and leave $b$ and $c$ unrestricted. The implementation of $A^{\text{Bayes}}_{\Gamma}$ and $A^{MN}_{\Gamma}$ requires the choice of a prior/weight function $\Gamma$ on this parameter space. We choose $\Gamma$ so that it puts all mass on models with $b = c = 0$, and with a uniform distribution on $d \in [-0.4, 1.4]$. Thus, we effectively construct the prediction sets that are, on average, as short as possible in the fractional model, and with guaranteed frequentist coverage in the unconstrained *bcd*-model. The uniform weighting over $d \in [-0.4, 1.4]$ makes the prediction sets informative over the whole range of anti-persistent ($d < 0$) to very persistent ($d \geq 1$) processes $x_t$.

As discussed in Section 2, the choice of $q$ may usefully thought of as a trade-off between efficiency and robustness. In principle, the central limit theorem for $(X'_{T1:q}, Y_T)'$ discussed in Section 3 and presented in the appendix holds for any fixed $q$, at least asymptotically. And the larger $q$, the smaller the (average) uncertainty about $Y_T$. This suggests that one should pick $q$ large to increase efficiency of the procedure.

At the same time, one might worry that approximations provided by the central limit theorem for $(X'_{T1:q}, Y_T)'$ become poor for large $q$. The concern is not only that the high-dimensional multivariate Gaussianity might fail to be an accurate approximation; more importantly, any parametric assumption about the shape of the local-to-zero spectrum becomes stronger for larger $q$. In particular, for a given sample size $T$, the assumption that the spectrum of $x$ over the frequencies $[-q\pi/T, q\pi/T]$ is well approximated by the spectrum of the *abd*-model becomes less plausible the larger $q$.

We are thus faced with a classic efficiency and robustness trade-off. Recall from the discussion of Figure 1 in Section 3, however, that the object of interest—the variability of long-run forecasts, as embodied by the conditional variance of $Y$ given $X_{1:q}$—is a low frequency

Table 1: Average Length of 90% $A_\Gamma^{MN}$ Prediction Sets

| | $r = 0.10$ | $r = 0.40$ | $r = 1.00$ |
|---|---|---|---|
| | (a) $\theta$ *unknown* | | |
| $q = 12$ | 1.00 | 1.00 | 1.00 |
| $q = 24$ | 0.92 | 0.92 | 0.88 |
| $q = 48$ | 0.89 | 0.87 | 0.79 |
| | (b) $\theta$ *known* | | |
| $q = 12$ | 0.90 | 0.79 | 0.63 |
| $q = 24$ | 0.87 | 0.77 | 0.62 |
| $q = 48$ | 0.86 | 0.76 | 0.61 |

Notes: This table shows the average asymptotic length of $A_\Gamma^{MN}$ prediction sets when $\theta \sim \Gamma$ and when the value of $\theta$ is unknown (panel (a)) and known (panel (b)) for selected values of the forecast horizon $r=h/T$. Average lengths are relative to the average value of $A_\Gamma^{MN}$ with $q = 12$ and $\theta$ unknown.

quantity that is essentially governed by properties of $x_t$ over frequencies $[-12\pi/T, 12\pi/T]$. Since the predictors $X_T(j)$ provide information for frequency $j\pi T$, this suggests that the marginal benefit of increasing $q$ beyond $q = 12$ is modest, at least with the spectrum known.

With the spectrum unknown, $X_{1:q}$ with larger $q$ provides additional information about its scale and its shape. The scale effect is most easily understood in the $I(0)$ model. As discussed above, the $I(0)$ prediction set is $\overline{x}_{1:T} \pm t^q_{(1-a/2)}(1 + r^{-1})^{1/2}T^{-1/2}s_{LR}$, where $s^2_{LR} = (T/q)X'_{T,1:q}X_{T,1:q}$. The average asymptotic length of this forecast is thus $2T^{-\alpha}t^q_{(1-a/2)}(1 + r^{-1})^{1/2}E\sqrt{X'_{1:q}X_{1:q}/q}$ with $X_{1:q} \sim \mathcal{N}(0, I_q)$, which decreases in $q$, since $t^q_{(1-a/2)}E\sqrt{X'_{1:q}X_{1:q}/q}$ is a decreasing function of $q$.[11] But the benefit of increasing $q$ is modest: for a 90% interval, the average length for $q \in \{24, 48, \infty\}$ is only $\{3.0\%, 4.4\%, 5.8\%\}$ shorter than for $q = 12$, for instance.

When the shape of the spectrum is unknown but parametrized, as in the *bcd*-model, increasing $q$ beyond 12 provides additional information about the shape of the spectrum over the crucial frequencies $[-12\pi/T, 12\pi/T]$. Table 1 quantifies the combined scale and shape effect by reporting the $\Gamma$-average of the expected length of forecasts of $Y/\sqrt{X'_{1:12}X_{1:12}}$ based on observations $X_{1:q}$ for $q \in \{12, 24, 48\}$. As a point of comparison, it also reports

[11] This is perfectly analogous to the wider confidence intervals that arise from the use of inconsistent HAC estimators; see Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002, 2005), and Müller (2012) for a review.

this average length for $\theta$ known, where reductions for larger $q$ are almost exclusively driven by the scale effect.[12] From panel (a), for $r = 0.40$, there is an 8% decrease in average length as $q$ increases from $q = 12$ to $q = 24$ and a further reduction of 5% for $q = 48$, where (from panel (b)) approximately one-fourth of the decrease comes from the unknown scale effect. Much of our empirical analysis uses 65 years of post-WWII data, so that a choice $q = 12$ relies on periodicities below 10.8 years, while $q = 24$ and $q = 48$ use periodicities below 5.4 and 2.7 years. The marginal benefit of increasing $q$ beyond $q = 12$ is modest, and $q$ must be chosen very much larger to substantially reduce forecast uncertainty overall. In our view, a concern about substantial spectral misspecification outweighs these potential gains, so that as a default, we suggest constructing the predictive sets with $q = 12$.

# 6 Prediction Sets for U.S. Macroeconomic Time Series

In this section we present prediction sets for eight U.S. economic time series for forecast horizons ranging from 10 to 75 years. These series include growth rates of per-capita values of real GDP and consumption, population, productivity, prices (as measured by the CPI and PCE deflator), and real stock returns. We construct prediction sets using post-WWII quarterly samples, and for several series, samples that extend into the early 20th century. We also examine prediction sets for inflation in Japan as a contrast to results for U.S. inflation. Sources and details of construction of the data are presented in the Data Appendix.

Prediction sets are constructed using the low-frequency transformations of the series with a benchmark value of $q = 12$, although we also investigate the robustness of these sets to larger and smaller values of $q$. We begin by discussing forecast intervals that arise in the $I(0)$ model, then turn to the empirical evidence for and impact of alternative $I(d)$ models, and finally present the Bayes and frequentist sets for the *bcd*-model discussed in Sections 4 and 5.

The data are shown Figure A.1 of the Appendix, which plots each time series together with its low-frequency component extracted by $X_{T,1:12}$, that is the series' projection on $\{\cos(j\pi(t-1/2)/T)\}_{j=0}^{12}$. Table 2 shows summary statistics and $I(0)$ prediction sets for the 25-year forecast horizon. As discussed in Section 4, the $I(0)$ Bayes and frequentist prediction sets are identical and equal to $\overline{x}_{1:T} \pm t^{12}_{(1-a/2)}(1+r^{-1})^{1/2}T^{-1/2}s_{LR}$ where $t^{12}_{(1-a/2)}$ is the $(1-\alpha)$th quantile of the Student-t distribution with 12 degrees of freedom, $s^2_{LR} = (T/q)X'_{T,1:12}X_{T,1:12}$

[12] These are for comparison only; the parameter of the *bcd*-model cannot be consistently estimated even as $q \to \infty$.

Table 2: Summary Statistics and I(0) Predictive Sets

| Series | $\overline{x}_{1:T}$ | $s_{LR}$ | Predictive Sets: I(0) Model 25-Year Horizon | | |
|---|---|---|---|---|---|
| | | | 50% | 80% | 90% |
| *Quarterly Post-WWII Series* | | | | | |
| GDP/Pop | 1.9 | 5.1 | (1.5, 2.3) | (1.1, 2.7) | (0.8, 3.0) |
| Cons/Pop | 2.1 | 4.7 | (1.7, 2.5) | (1.3, 2.8) | (1.1, 3.1) |
| TF Prod | 1.3 | 4.0 | (1.0, 1.6) | (0.7, 1.9) | (0.5, 2.1) |
| Labor Prod | 2.2 | 3.7 | (1.9, 2.5) | (1.6, 2.8) | (1.4, 3.0) |
| Population | 1.2 | 1.5 | (1.1, 1.3) | (1.0, 1.4) | (0.9, 1.5) |
| Inflation (PCE) | 3.3 | 9.0 | (2.5, 4.0) | (1.8, 4.7) | (1.4, 5.2) |
| Inflation (CPI) | 3.6 | 10.4 | (2.8, 4.5) | (2.0, 5.3) | (1.5, 5.8) |
| Inflation (CPI, Japan, from 1960) | 3.2 | 15.0 | (1.9, 4.5) | (0.7, 5.7) | (0.0, 6.5) |
| Stock Returns | 6.7 | 30.1 | (4.3,9.2) | (1.9,11.5) | (0.4,13.1) |
| *Longer Span Data Series* | | | | | |
| GDP/Pop (Annual, 1901-2011) | 1.9 | 6.0 | (1.0, 2.9) | (0.1, 3.7) | (−0.4, 4.3) |
| Cons/Pop (Annual, 1901-2011) | 1.7 | 3.4 | (1.2, 2.3) | (0.7, 2.8) | (0.4, 3.1) |
| Population (Annual, 1901-2011) | 1.3 | 1.1 | (1.1, 1.4) | (0.9, 1.6) | (0.8, 1.7) |
| Inflation (CPI) (Annual, 1914-2011) | 3.2 | 8.6 | (1.8, 4.5) | (0.6, 5.8) | (−0.3, 6.6) |
| Stock Returns (Quarterly, 1926:Q2-2012:Q1) | 6.4 | 29.3 | (4.0,8.7) | (1.8, 10.9 | (0.4, 12.3) |

Notes: This table shows in-sample mean ( $\overline{x}_{1:T}$ ), estimated long-run standard deviation ($s_{LR}$), and prediction sets computed for the *I*(*0*) model. The Bayes and frequentist prediction sets coincide in the *I*(0) model and are given by $\overline{x}_{1:T} \pm t^{12}_{(1-\alpha/2)}(1+r^{-1})^{1/2}T^{-1/2}s_{LR}$, where $t^{12}_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of the *t*-distribution with 12 degrees of freedom and $r = h/T$ is the forecast horizon expressed a fraction of the sample size.

is an estimate of the long-run variance, and $r = h/T$ is forecast horizon expressed as a fraction of the sample size. For example, from the first row of the table, the growth rate of real per-capita GDP averaged 1.9% (at an annual rate) over the 1947:Q2-2012:Q2 sample period, with an estimated long-run standard deviation of $s_{LR} = 5.1\%$. Over the next 25 years per-capita GDP is forecast to grow at an average rate between 1.5% and 2.3% with probability 50%; that is the 50% prediction set is $(1.5, 2.3)$. The 80% and 90% prediction sets are wider, $(1.1, 2.7)$ and $(0.8, 3.0)$, respectively.

The $I(0)$ prediction sets are based on the strong assumption that the spectrum is perfectly flat over the relevant low frequencies, ruling out any persistence, stochastic trends or other low-frequency dynamics. This is especially implausible for series like CPI inflation. In the U.S., CPI inflation averaged 3.6% over the sample period with a long-run standard deviation of 10.4%. This yields a 25-year ahead 50% prediction set of $(2.8, 4.5)$ in the $I(0)$ model. This suggests that over the next 25 years the U.S. will experience significantly higher average values of inflation than those experienced over the past decade and well above the Federal Reserve's inflation target. Perhaps even more problematic is the 50% prediction interval for Japanese inflation of $(1.9, 4.5)$, a range comfortably above the very low levels of inflation that Japan has experienced over the past two decades. Arguably, one problem with these $I(0)$ prediction sets is that they are centered around the in-sample means for for inflation (3.6% for U.S. CPI inflation and 3.2% for Japanese inflation). However, inflation is very persistent (Figure A.1, panels g and h), so that this centering is problematic. The remaining tables and figures study the persistence properties of the various time series and present prediction sets for processes that are more general than the $I(0)$ model.

Table 3 summarizes the persistence properties of the time series in terms of the value of $d$ in the $I(d)$ model. It shows the log-likelihood values for various values of $d$, where the log-likelihood is based on $X^s_{T,1:12}$ and its asymptotic distribution (8), and the log-likelihood of the $I(0)$ model is normalized to zero. The post-WWII U.S. inflation data yields log-likelihoods with maxima around $d = 0.6$ with corresponding log-likelihood values that are between 1.9 and 2.9 times larger than the $I(0)$ model. The log-likelihood values for inflation suggests two conclusions: first, inflation is more persistent than the $I(0)$ model (so that the $I(0)$ prediction intervals in Table 2 are likely to be invalid), and second, there is considerable uncertainty about the exact degree of persistence. For example, the MLE of $d$ is approximately $\widehat{d}^{MLE} \approx 0.7$ for post-war U.S. PCE inflation, but values of $d$ ranging from $d = 0.3$ to $d = 1.1$ reduce the log-likelihood by less than 1. This reflects the limited long-run information in a 65-year sample. Looking at post-WWII real per-capita GDP suggests that persistence is not large

Table 3: Log Likelihood Values for I(*d*) model

| Series | *d* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | -0.4 | -0.2 | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| *Quarterly Post-WWII Series* | | | | | | | | | | |
| GDP/Pop | 0.5 | 0.4 | 0.0 | -0.6 | -1.6 | -2.8 | -4.3 | -6.0 | -8.1 | -10.6 |
| Cons/Pop | 0.3 | 0.2 | 0.0 | -0.4 | -1.0 | -1.8 | -3.0 | -4.4 | -6.2 | -8.4 |
| TF Prod | -2.6 | -1.1 | 0.0 | 0.6 | 0.7 | 0.1 | -0.9 | -2.4 | -4.2 | -6.6 |
| Labor Prod | -1.5 | -0.6 | 0.0 | 0.2 | 0.1 | -0.3 | -1.0 | -2.0 | -3.4 | -5.3 |
| Population | -4.0 | -2.1 | 0.0 | 2.1 | 4.1 | 5.7 | 6.8 | 7.2 | 6.8 | 5.4 |
| Inflation (PCE) | -3.5 | -1.6 | 0.0 | 1.3 | 2.3 | 2.9 | 2.9 | 2.5 | 1.6 | 0.0 |
| Inflation (CPI) | -3.3 | -1.5 | 0.0 | 1.1 | 1.7 | 1.9 | 1.5 | 0.7 | -0.5 | -2.3 |
| Infl. (CPI,Japan) | -5.0 | -2.4 | 0.0 | 2.0 | 3.5 | 4.1 | 3.9 | 3.0 | 1.7 | -0.3 |
| Stock Returns | -1.6 | -0.6 | 0.0 | 0.2 | 0.0 | -0.6 | -1.5 | -2.8 | -4.4 | -6.4 |
| *Longer Span Data Series* | | | | | | | | | | |
| GDP/Pop | 1.0 | 0.7 | 0.0 | -1.0 | -2.2 | -3.6 | -5.2 | -6.9 | -8.8 | -11.2 |
| Cons/Pop | -2.5 | -0.9 | 0.0 | 0.1 | -0.3 | -1.1 | -2.2 | -3.5 | -5.0 | -6.9 |
| Population | -1.6 | -0.8 | 0.0 | 0.8 | 1.4 | 1.8 | 1.9 | 1.7 | 1.1 | -0.1 |
| Inflation (CPI) | -1.6 | -0.5 | 0.0 | 0.2 | 0.1 | -0.1 | -0.4 | -0.8 | -1.5 | -2.6 |
| Stock Returns | -0.6 | -0.1 | 0.0 | -0.2 | -0.8 | -1.7 | -2.8 | -4.2 | -5.9 | -8.0 |

Notes: This table shows the log-likelihood of *d* in the *I*(*d*) model based on the large-sample density of $X^{s}_{T,1:12}$ given in equation (7).

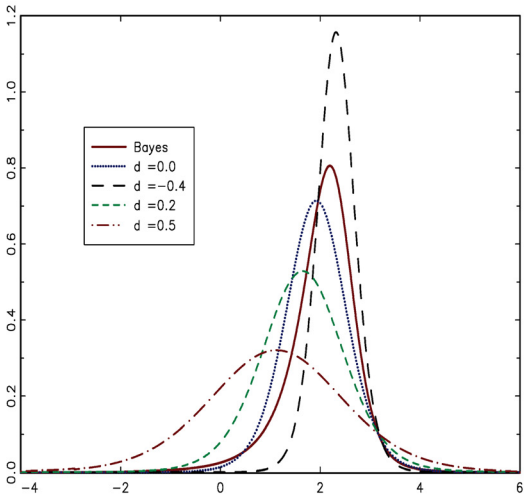
for this series (values of $d > 0.6$ yield a log-likelihood 3 points lower than the $I(0)$ model), but values of $d$ ranging from $-0.4$ (suggesting some reversion to a linear trend in the level of GDP, so that the growth rate is overdifferenced) to 0.2 (slight persistence in the GDP growth rates) all fit the data reasonably well.

Figure 3 shows how uncertainty about $d$ translates into prediction uncertainty. It shows 25-year ahead predictive densities constructed using four different values of $d$ for four of the series listed in Table 2. Panel (a) shows results for real per-capita GDP, and the predictive densities are shown for $d = -0.4$, 0.0, 0.2 and 0.5, a plausible range of values from Table 3. As $d$ increases, the variance of the predictive density increases (because more persistence leads to larger variability in future average growth) and the mode of the density shifts to the left (reflecting the persistent effect of the slow growth experienced at the end of sample). Also plotted in the figure is the Bayes predictive density constructed using a flat prior on $-0.4 \leq d \leq 1.4$. For real GDP, because (from Table 3) negative values of $d$ fit the data somewhat better than the $I(0)$ model, the Bayes predictive density is shifted to the right of the $I(0)$ density, reflecting low-frequency anti-persistence in the forecast period following a decade of lower-than average growth at the end of the sample. Table 3 suggests that total factor productivity may be somewhat more persistent than the $I(0)$ model and this is reflected in panel (b) of Figure 3 in a Bayes predictive density that is more disperse than the $I(0)$ density; the Bayes density is very close to the density computed using $d = 0.4$. Panel (c) shows results for CPI inflation, where the Bayes density is much more disperse than the $I(0)$ model, but much less disperse than the $I(1)$ model. Finally, panel (d) of Figure 3 shows results for post-WWII real stock returns. From Table 3, real returns may be somewhat more persistent than an $I(0)$ process, and thus the Bayes predictive density in Figure 3 is more disperse than the $I(0)$ density.

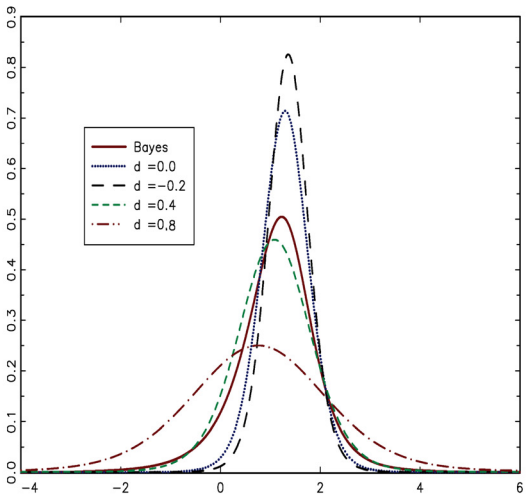
Table 4 and Figure A.2 show prediction sets for other horizons. These show equal-tail Bayes prediction sets constructed using the $I(d)$ model with flat prior for $d \in [-0.4, 1.4]$ together with $A_{\Gamma}^{MN}$ frequentist prediction sets that have guaranteed coverage in the *abd*-model, as discussed in Section 4. Figure A.2 plots prediction sets for horizons ranging from 10 to 75 years, and Table 4 shows the sets for four forecast horizons, 10, 25, 50, and 75 years. The prediction sets were computed separately for each horizon, so coverage is pointwise in the forecast horizon. In Figure A.2 the boundary of the Bayes sets are plotted as solid curves and the frequentist sets as dashed curves. In many cases the sets coincide, so that only one curve is visible in the plot. In Table 4 the Bayes sets are shown in parentheses, and the corresponding frequentist set is shown in brackets when it differs from the Bayes

Figure 3: 25-Year Ahead Predictive Densities

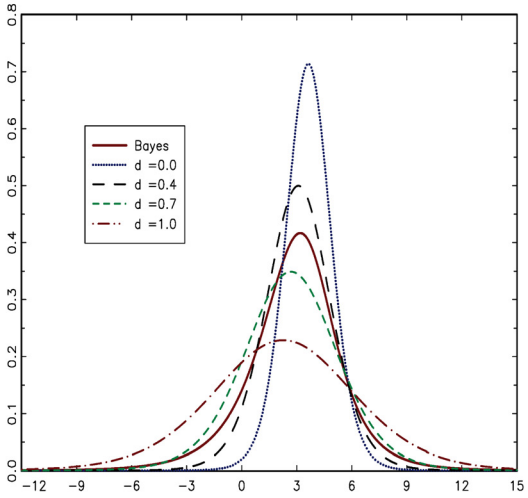
A. Real Per-Capita GDP




B. Total Factor Productivity




C. Inflation (CPI)




D. Stock Returns

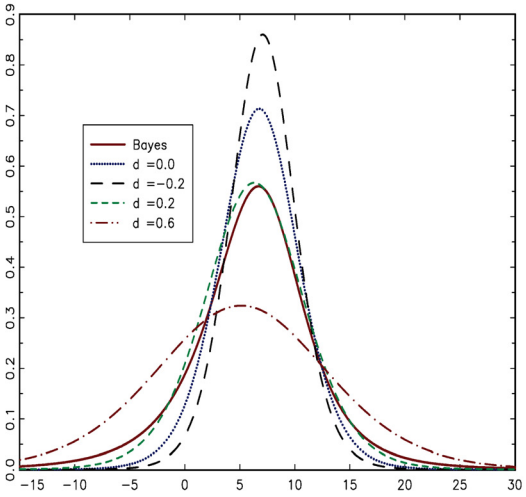


Table 4: Prediction Sets

a. 50% Coverage

| Series | Forecast horizon (in years) | | | |
|---|---|---|---|---|
| | **10** | **25** | **50** | **75** |
| *Quarterly Post-WWII Series* | | | | |
| GDP/Pop | (1.6, 2.9) | (1.7, 2.5) | (1.8, 2.3) | (1.8, 2.2) |
| Cons/Pop | (1.6, 2.9) | (1.8, 2.6) | (1.8, 2.4) | (1.9, 2.4) |
| TF Prod | (0.5, 1.6)<br>[0.2, 1.6] | (0.7, 1.6)<br>[-0.1, 1.6] | (0.8, 1.6)<br>[-0.4, 1.8] | (0.8, 1.6)<br>[-0.7, 2.0] |
| Labor Prod | (1.5, 2.6) | (1.7, 2.5) | (1.7, 2.5)<br>[1.4, 2.5] | (1.8, 2.5)<br>[1.1, 2.6] |
| Population | (0.6, 1.0) | (0.6, 1.0) | (0.5, 1.1) | (0.4, 1.2) |
| Inflation (PCE) | (1.0, 3.5)<br>[1.0, 4.0] | (0.9, 3.8) | (0.9, 4.1) | (0.8, 4.2) |
| Inflation (CPI) | (1.3, 4.2)<br>[1.3, 4.5] | (1.4, 4.4) | (1.5, 4.6) | (1.4, 4.6) |
| Infl. (CPI,Japan) | (-1.9, 2.4) | (-1.8, 3.1)<br>[-2.7, 3.1] | (-1.7, 3.8)<br>[-3.8, 3.8] | (-1.7, 4.1)<br>[-4.6, 4.1] |
| Stock Returns | (2.0,10.2) | (3.0, 9.5) | (3.6, 9.1) | (3.7, 9.0) |
| *Longer Span Data Series* | | | | |
| GDP/Pop | (1.0, 3.7) | (1.4, 2.9) | (1.5, 2.6) | (1.6, 2.5) |
| Cons/Pop | (0.6, 2.3) | (0.9, 2.2) | (1.1, 2.1)<br>[1.1, 2.5] | (1.2, 2.1)<br>[1.1, 2.8] |
| Population | (0.6, 1.1) | (0.6, 1.2) | (0.6, 1.3) | (0.6, 1.3) |
| Inflation (CPI) | (0.8, 5.1) | (1.2, 4.9) | (1.4, 4.8) | (1.5, 4.8) |
| Stock Returns | (2.9,10.2) | (3.9, 8.8) | (4.4, 8.3) | (4.7, 8.0) |

Table 4: Prediction Sets
(Continued)

b. 80% Coverage

| Series | Forecast horizon (in years) | | | |
|---|---|---|---|---|
| | **10** | **25** | **50** | **75** |
| *Quarterly Post-WWII Series* | | | | |
| GDP/Pop | (0.9, 3.4) | (1.2, 2.8)<br>[0.4 2.8] | (1.4, 2.5)<br>[-0.4, 2.5] | (1.4, 2.5)<br>[-0.9, 2.5] |
| Cons/Pop | (0.8, 3.4) | (1.2, 2.9)<br>[0.7, 2.9] | (1.3, 2.7)<br>[0.1, 2.7] | (1.4, 2.6)<br>[-0.4, 2.6] |
| TF Prod | (-0.1, 2.1)<br>[-0.5, 2.1] | (0.1, 2.0)<br>[-0.7, 2.3] | (0.2, 2.1)<br>[-1.5, 2.8] | (0.2, 2.1)<br>[-2.2, 3.4] |
| Labor Prod | (1.0, 3.0) | (1.1, 2.9)<br>[0.9, 3.1] | (1.2, 2.9)<br>[0.4, 3.4] | (1.2, 3.0)<br>[0.0, 3.8] |
| Population | (0.5, 1.1) | (0.3, 1.2)<br>[0.3, 1.3] | (0.1, 1.4) | (-0.1, 1.5)<br>[-0.3, 1.5] |
| Inflation (PCE) | (-0.2, 4.7)<br>[-0.2, 4.8] | (-0.7, 5.2) | (-1.3, 5.8) | (-1.8, 6.3) |
| Inflation (CPI) | (-0.1, 5.6) | (-0.4, 5.9) | (-0.7, 6.3) | (-0.9, 6.6) |
| Infl. (CPI,Japan) | (-4.2, 4.3)<br>[-4.6, 4.7] | (-4.7, 5.3)<br>[-6.6, 5.5] | (-5.3, 6.3)<br>[-9.8, 7.0] | (-5.8, 7.0)<br>[-12.5, 8.8] |
| Stock Returns | (-2.2,13.9) | (-0.9,12.7)<br>[-1.1, 12.8] | (-0.3, 12.3)<br>[-3.8, 12.3] | (-0.2, 12.3)<br>[-6.6, 13.4] |
| *Longer Span Data Series* | | | | |
| GDP/Pop | (-0.4, 5.0)<br>[-0.4, 5.2] | (0.5, 3.7) | (0.9, 3.1) | (1.1, 2.9) |
| Cons/Pop | (-0.2, 3.1)<br>[-0.2, 3.2] | (0.2, 2.8)<br>[0.2, 3.1] | (0.4, 2.7)<br>[0.0, 3.4] | (0.5, 2.6)<br>[-0.5, 3.7] |
| Population | (0.4, 1.4) | (0.3, 1.5) | (0.2, 1.6) | (0.1, 1.6) |
| Inflation (CPI) | (-1.3, 7.1) | (-1.0, 6.9) | (-1.1, 7.1) | (-1.2, 7.2) |
| Stock Returns | (-0.6, 13.7) | (1.4, 11.4) | (2.1, 10.3) | (2.5, 10.1) |

Table 4: Prediction Sets
(Continued)

c. 90% Coverage

| Series | Forecast horizon (in years) | | | |
|---|---|---|---|---|
| | **10** | **25** | **50** | **75** |
| *Quarterly Post-WWII Series* | | | | |
| GDP/Pop | (0.4, 3.7)<br>[0.2, 3.7] | (0.8, 3.0)<br>[-0.4, 3.0] | (1.0, 2.7)<br>[-1.2, 2.7] | (1.1, 2.6)<br>[-1.8, 2.9] |
| Cons/Pop | (0.3, 3.8) | (0.7, 3.1)<br>[-0.2, 3.1] | (0.8, 2.9)<br>[-0.9, 2.9] | (0.9, 2.8)<br>[-1.5, 3.1] |
| TF Prod | (-0.4, 2.4)<br>[-0.9, 2.5] | (-0.3, 2.4)<br>[-1.3, 2.8] | (-0.3, 2.4)<br>[-2.4, 3.6] | (-0.3, 2.5)<br>[-3.4, 4.5] |
| Labor Prod | (0.6, 3.4) | (0.7, 3.3)<br>[0.4, 3.5] | (0.7, 3.3)<br>[-0.4, 4.1] | (0.7, 3.4)<br>[-1.1, 4.7] |
| Population | (0.4, 1.2) | (0.1, 1.4) | (-0.2, 1.6)<br>[-0.5, 1.7] | (-0.5, 1.7)<br>[-0.9, 2.0] |
| Inflation (PCE) | (-1.0, 5.4) | (-1.9, 6.2) | (-3.1, 7.3) | (-4.1, 8.2) |
| Inflation (CPI) | (-1.1, 6.4) | (-1.6, 7.0)<br>[-1.8, 7.0] | (-2.4, 7.7)<br>[-3.4, 7.7] | (-3.0, 8.4)<br>[-5.2, 8.6] |
| Infl. (CPI,Japan) | (-5.6, 5.6)<br>[-6.4, 5.9] | (-6.8, 6.8)<br>[-9.2, 7.7] | (-8.3, 8.4)<br>[-14.6, 11.6] | (-9.4, 9.5)<br>[-18.8, 15.3] |
| Stock Returns | (-5.0, 16.4) | (-4.0, 15.3)<br>[-6.0, 15.3] | (-3.7, 15.2)<br>[-10.5, 17.0] | (-3.8, 15.4)<br>[-14.9, 20.4] |
| *Longer Span Data Series* | | | | |
| GDP/Pop | (-1.3, 5.9)<br>[-1.4, 6.0] | (-0.1, 4.1) | (0.5, 3.5) | (0.7, 3.2) |
| Cons/Pop | (-0.8, 3.6)<br>[-0.8, 3.8] | (-0.4, 3.2)<br>[-0.4, 3.5] | (-0.1, 3.0)<br>[-1.0, 4.0] | (-0.1, 3.0)<br>[-1.6, 4.4] |
| Population | (0.2, 1.6) | (0.1, 1.6) | (-0.2, 1.7) | (-0.4, 1.9) |
| Inflation (CPI) | (-2.6, 8.5) | (-2.6, 8.4) | (-3.4, 9.1) | (-4.0, 9.8) |
| Stock Returns | (-2.9, 15.8) | (-0.7, 13.1) | (0.1, 12.2)<br>[-1.0, 12.2] | (0.5, 12.0)<br>[-2.6, 12.0] |

Notes: This table shows the 50%, 80% and 90% prediction sets for forecast horizons, $h = 10$, 25, 50, and 75 years. The Bayes sets are shown in parentheses and are based on the $I(d)$ model with uniform prior for $-0.4 \leq d \leq 1.4$. The *MN*-frequentist sets are shown in brackets, and are omitted if they coincide with the Bayes sets. By construction the *MN*-frequentist sets control asymptotic coverage in the *bcd*-model with $-0.4 \leq d \leq 1.4$, and $b$ and $c$ unrestricted.

set. Results are shown for 50%, 80%, and 90% prediction sets.

We now discuss the results for specific series in more detail.

*Real per capita GDP.* The Bayes predictive sets for per-capita GDP narrow as the forecast horizon increases, consistent with the reduction in variance of the sample mean for an $I(0)$ process. At the 75-year horizon the 80% Bayes prediction interval is 1.4 to 2.5, which roughly coincides with the interval reported by the Congressional Budget Office (2005) for 75-year forecasts beginning in 2004. The coincidence of the Bayes/CBO sets arises despite important differences in the way they are computed. The CBO interval is based on simulations computed from its long-run model with inputs such as TFP growth simulated from estimated $I(0)$ models. The CBO interval differs from the Bayes interval in two important respects. First, because the simulations are carried out using fixed values of the model parameters, the CBO method ignores the parameter uncertainty in $\overline{x}_{1:T}$ (as an estimate of $\mu$) and $s^2_{LR}$ (as a an estimate of the long-run variance). Ignoring this uncertainty leads the CBO interval to underestimate uncertainty in the predictions. Second, in the CBO model, GDP growth is $I(0)$, while the Bayes method allows values of $d$ that differ from $d = 0$. The results in Table 3 and Figure 3 suggest that GDP growth is plausibly characterized by a process with some low-frequency anti-persistence, and this translates into less forecast uncertainty than the CBO's $I(0)$ model. Thus, the CBO method tends to understate forecast uncertainty because it ignores parameter uncertainty in the estimated mean and long-run variance, and to overstate forecast uncertainty because its model does not capture long-run anti-persistence associated with negative values of $d$. Apparently these two errors cancel, so that the CBO prediction interval coincides with the Bayes set.

The frequentist set coincides with the Bayes set for 50% coverage and for (relatively) short horizons for higher coverage. However, at longer horizons the 80% and 90% frequentist sets differ from Bayes sets and include smaller values of average GDP growth rates. Apparently, to guarantee high coverage uniformly in the *abd*-model at long horizons, the frequentist sets allow for the possibility of more persistence in the GDP process, so that the slow-growth rates of the past decade are predicted to potentially persist into the future. A comparison of the prediction sets constructed using the post-WWII data and the long-annual (1901-2011) series shows that the pre-WWII data tend to widen the predictions sets, presumably because of the higher (long-run) variance in the pre-WWII data evident in Figure A.1.

*Productivity.* Table 3 indicates that productivity (TFP and average labor productivity) may have somewhat greater than $I(0)$ persistence. This translates into prediction sets that are wider than $I(0)$ sets (see Figure 3 for a comparison of $I(0)$ and Bayes predictive densities

for TFP), particularly for frequentist sets at large forecast horizons. Figure A.2 shows that the Bayes intervals are essentially flat as the forecast horizon increases (unlike in an $I(0)$ model, where the intervals narrow), while the frequentist sets widen (the unmodified Bayes intervals systematically undercover for larger values of $d$, forcing the frequentist intervals to more heavily weigh the possibility of larger $d$).

*Population.* Figure A.1 shows considerable low-frequency variability in U.S. population growth over the 20th century and the post-WWII period.[13] Immigration and fertility dynamics are presumably at the source of these long swings. Table 3 indicates that $\widehat{d}^{MLE}$ is very close to unity over both sample periods, with the $I(1)$ likelihood more than 7-log points higher than in the $I(0)$ model. Figure A.2 shows prediction intervals that widen as the forecast horizon increases, a familiar characteristic of $I(1)$ predictive densities. Frequentist and Bayes prediction sets are similar, as the Bayes sets already put sufficiently large weight on large values of $d$. There is little difference in the sets constructed using the post-WWII samples and long-samples.

*Inflation.* As discussed above, the inflation process is characterized by more than $I(0)$ persistence, and this is reflected in the prediction sets in two ways. First, they are not centered at the sample mean of the series, but rather at a level dictated by the values near the end of sample period, and second, the prediction sets widen with the forecast horizon. The prediction intervals indicate considerable uncertainty in inflation even at relatively short horizons; this is true for Bayes and frequentist sets (which essentially coincide). For example, while the 10-year 50% Bayes predictive set for U.S. CPI inflation is $(1.3, 4.2)$, the 80% set widens to $(-0.1, 5.6)$, and the 90% set widens further to $(-1.1, 6.5)$.

These predictions sets may strike some readers as too large, but it is instructive to consider the history of Japan where the 10-year moving average of CPI inflation was less than zero from 2003 through the end of the sample. Moreover, they are in line with predictive densities derived from asset prices. For example, Kitsul and Wright (2012) use CPI-based derivatives to compute market-based risk-neutral predictive densities for 10-year ahead average values of inflation. They find deflation (average inflation less than 0%) probabilities that averaged approximately 15% over 2011 and "high inflation" (average inflation greater than 4%) of 30%.[14] The corresponding probabilities computed from the Bayes predictive density constructed using the post-WWII data are 11% for deflation and 28% for high inflation.

[13] The quaterly post-WWII population series is not seasonally adjusted and shows substantial seasonal variation. Because we focus on low-frequency transformations of the series, this seasonality does not affect the empirical results.

[14] See Kitsul and Wright (2012), Figures 3 and 4.

*Stock Returns.* Table 3 indicates that post-WWII real stock returns exhibit slightly more persistence than is implied by the $I(0)$ model, and this translates into prediction sets that are wider than implied by the $I(0)$ model. For example, at the 25-year horizon, the $I(0)$ 80% prediction set (from Table 2) is $(1.9, 11.5)$ while the corresponding Bayes prediction set (from Table 4) is $(-0.9, 12.7)$ and the frequentist set is wider still $(-1.1, 12.8)$. The longer-span data suggest somewhat less persistence ($\widehat{d}^{MLE} = 0.0$ for the 1926-2001 sample) yielding Bayes and frequentist prediction intervals that are somewhat narrower than those constructed using the post-WWII data.

Pastor and Stambaugh (2012) survey the large literature on long-run stock return volatility and construct Bayes predictive densities using models that allow for potentially persistent components in returns and incorporate parameter uncertainty. While their results rely on more parametric models than ours—they use all frequencies and exact Gaussian likelihoods—our empirical conclusions are similar. Using our notation, Pastor and Stambaugh (2012) are concerned with the behavior of the variance of $\sqrt{h}\bar{x}_{T+1:T+h}$ and how this variance changes with the forecast horizon $h$. If the variance of $\sqrt{h}\bar{x}_{T+1:T+h}$ is unchanged as $h$ increases, and if the predictive density is Gaussian, then the width of prediction intervals for $\bar{x}_{T+1:T+h}$ will be proportional to $h^{-1/2}$. Pastor and Stambaugh find that the variance of $\sqrt{h}\bar{x}_{T+1:T+h}$ is not constant, but rather increases with $h$. In our results,the Bayes predictive sets narrow as $h$ increases as $h$ increases, but more slowly than $h^{-1/2}$, consistent with the Pastor-Stambaugh findings.

*Results for different values of q.* As discussed in Sections 3 and 5, the choice of $q = 12$ involved an efficiency/robustness trade-off, where a larger value of $q$ results in more information about the scale and shape parameter, but potential misspecification because the higher-frequency spectrum may not be well-described by the same model and parameter. It is therefore interesting to see how the prediction sets vary with $q$, and this is reported in Table 5, which shows the 80% prediction sets for the 25-year ahead forecasts for $q = 6, 9, 12, 24$, and 48. Looking across all of the entries, the prediction sets behave roughly as expected, in the sense that they remain centered at roughly the same value but tend to narrow as $q$ increases. For example, averaging across the 14 series, the $q = 48$ prediction set is 2% narrower than the $q = 12$ prediction set. There are two noteworthy exceptions. First, both per-capita GDP and consumption (quarterly post-WWII) show a lower bound for the $q = 6$ prediction sets that is more than one percentage point lower than the corresponding value for the $q = 12$ sets. This arises because the $q = 6$ log-likelihood functions for $d$ are much flatter than the $q = 12$ functions summarized in Table 3. Thus, when $q = 6$, the prediction sets account for

Table 5: 80% Predictive Sets for Various Values of $q$
25-Year Horizon

| Series | $q = 6$ | $q = 9$ | $q = 12$ | $q = 24$ | $q = 48$ |
|---|---|---|---|---|---|
| | *Quarterly Post-WWII Data* | | | | |
| GDP/Pop | (-0.4, 2.5) | (0.5, 2.7)<br>[ -0.1, 2.7] | (1.2, 2.8)<br>[0.4 2.8] | (1.3, 2.7)<br>[ 0.6, 2.7] | (1.1, 2.7)<br>[ 0.1, 2.7] |
| Cons/Pop | (0.1, 2.8) | (0.9, 2.9)<br>[0.7, 2.9] | (1.2, 2.9)<br>[0.7, 2.9] | (1.3, 2.8)<br>[ 0.5, 2.8] | (1.2, 2.8)<br>[ 0.2, 2.8] |
| TF Prod | (-0.3, 1.8)<br>[-0.5, 1.9] | (-0.7, 1.8)<br>[-0.8, 2.1] | (0.1, 2.0)<br>[-0.7, 2.3] | (0.4, 2.0)<br>[-0.9, 2.5] | (0.5, 2.0)<br>[-1.0, 2.6] |
| Labor Prod | (0.9, 2.9) | (0.8, 2.8)<br>[0.8, 3.0] | (1.1, 2.9)<br>[0.9, 3.1] | (1.3, 2.8)<br>[ 0.7, 3.1] | (1.4, 2.8)<br>[ 0.7, 3.2] |
| Population | (0.2, 1.3)<br>[0.2, 1.4] | (0.4, 1.3)<br>[0.3, 1.4] | (0.3, 1.2)<br>[0.3, 1.3] | (0.1, 1.1) | (0.3, 1.1) |
| Inflation (PCE) | (-0.4, 5.6) | (-0.8, 5.2) | (-0.7, 5.2) | (0.4, 5.0)<br>[-1.0, 5.3] | (0.5, 4.8)<br>[-0.9, 5.6] |
| Inflation (CPI) | (-0.5, 6.2) | (-0.8, 5.9) | (-0.4, 5.9) | (0.7, 5.6)<br>[-1.0, 5.9] | (0.5, 5.7)<br>[-0.3, 5.7] |
| Infl. (CPI,Japan) | (-3.9, 6.4)<br>[-7.2, 6.9] | (-4.9, 5.7)<br>[-6.4, 5.7] | (-4.7, 5.3)<br>[-6.6, 5.5] | (-3.6, 5.5)<br>[-4.5, 5.5] | (-3.5, 5.3) |
| Stock Returns | (-3.0, 12.8) | (-0.3, 13.4)<br>[-0.6, 13.4] | (-0.9,12.7)<br>[-1.1, 12.8] | (1.2 ,12.3)<br>[-3.0 ,12.5] | (1.5 ,11.8)<br>[-2.2 ,12.2] |
| | *Longer Span Data Series* | | | | |
| GDP/Pop | (0.3, 3.3) | (0.5, 3.4) | (0.5, 3.7)<br>[0.3, 3.8] | (0.4, 3.5) | (0.3, 3.6) |
| Cons/Pop | (0.3, 2.9) | (-0.1, 2.7)<br>[-0.1, 2.9] | (0.2, 2.8)<br>[0.1, 2.9] | (0.6, 2.9)<br>[ 0.5, 3.3] | (0.6, 2.9)<br>[0.3, 3.3] |
| Population | (0.6, 1.8) | (0.5, 1.6) | (0.3, 1.5)<br>[0.1, 1.6] | (0.3, 1.4) | (0.1, 1.3) |
| Inflation (CPI) | (-0.7, 7.0) | (-0.5, 6.6) | (-1.0, 6.9)<br>[-1.5, 7.8] | (0.0, 6.0) | (-0.6, 6.5) |
| Stock Returns | (0.0 ,12.1) | (-0.8, 11.1) | (1.4, 11.4)<br>[-1.2, 14.0] | (2.7, 10.2) | (1.9, 10.9) |

Notes: Bayes sets are shown in parentheses and *MN*-frequentist sets are shown in brackets when they differ from Bayes sets. See notes to Table 3 for additional information.

the possibility of more persistence in the series and this includes an extrapolation of the low-growth levels experienced by the U.S. economy over the past decade. The second exception is quarterly U.S. inflation. Here the lower bounds of the $q = 24$ and $q = 48$ prediction sets are more than one percentage point higher than for $q = 12$. This arises because the $q = 24, 48$ likelihood values suggest less persistence than $q = 12$ value. (When $q = 24, 48$ the likelihood peaks at values of $d$ around 0.4 while the peak is around $d = 0.7$ for $q = 12$.)

# 7 Additional Remarks

We conclude by discussing two issues. First, we analyze the robustness of our prediction sets relative to data generating processes with pronounced second moment variability, and then we briefly discuss the challenges of extending our framework to a multivariate setting.

## 7.1 Heteroskedasticity

The large-sample results presented in Section 3 allow for processes with conditional heteroskedasticity that may be persistent, but that eventually vanishes in the sense of assumption (i) of Theorem 1 in the Appendix. That said, these large-sample results may provide poor approximations in samples of the size considered in the last section (65 years of quarterly data) for series with persistent changes in second moments. We investigate this in this subsection with a simulation experiment focusing on the $I(0)$ model, which approximately describes stock returns, and the local-level model, which approximately describes the U.S. inflation rate. The $I(0)$ model with stochastic volatility has the form

$$x_t = \sigma_t \varepsilon_t \tag{13}$$

where

$$\ln(\sigma_t^2) = \gamma(1-\phi) + \phi \ln(\sigma_{t-1}^2) + e_t \tag{14}$$

and where $\{\varepsilon_t\}$ and $\{e_t\}$ are independent with $\varepsilon_t \sim iid\mathcal{N}(0,1)$ and $e_t \sim iid\mathcal{N}(0,\sigma_e^2)$. For the simulations we calibrate $\sigma_e^2$ using the quarterly post-war stock returns from Section 6, the value of $\gamma$ does not matter because of scale invariance, and we investigate coverage for several values of $\phi$, keeping the unconditional variance of $\ln(\sigma_t^2)$ fixed at the sample analogue. The local-level model with stochastic volatility has the form

$$x_t = \varepsilon_{1t} + (bT)^{-1} \sum_{s=1}^{t} \sigma_s \varepsilon_{2s}$$

where $\ln(\sigma_t)$ follows the AR(1) process in (14) and where $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ are mutually independent and independent of $\{e_t\}$ and distributed $\varepsilon_{it} \sim iidN(0, \sigma_\varepsilon^2)$, $i = 1, 2$. In this local-level model, stochastic volatility is present in the $I(1)$ component but not in the $I(0)$ component, consistent with the evidence for U.S. inflation provided in Stock and Watson (2007). The values of $\sigma_\varepsilon$, $b$, $\gamma$, and $\sigma_e$ are calibrated using the estimated inflation stochastic volatility estimated for CPI inflation from Stock and Watson (2010), and simulations are carried out for several values of $\phi$ (again keeping the unconditional variance of $\ln(\sigma_t^2)$ fixed at a sample estimate). In the simulations we set $T = 240$ to correspond with our post-WWII empirical analysis.

The results are shown in Table 6. Each row of the table corresponds to a different forecast horizon, $r$, and the various columns correspond to different values of $\phi$ over four sets of experiments. In the first experiment $\phi = 0$, so volatility is white noise, resulting in non-normal i.i.d. disturbances. Coverage rates for 90% Bayes and MN-frequentist prediction sets are shown in the second column of the table. The Bayes prediction sets tend to over-cover in the $I(0)$ model and under-cover in the local-level model. (Recall that the Bayes sets have asymptotic 90% coverage by construction in the $I(d)$ model when $d \sim U[-0.4, 1.4]$). In contrast, the MN-frequentist sets over-cover in the $I(0)$ model when $r$ is large, but have coverage of nearly 90% in the local-level model. (Recall that the MN-frequentist have asymptotic coverage of at least 90% by construction for $-0.4 \leq d \leq 1.4$ and all values of $a$ and $b$.) The next three columns of the table show coverage rates for $\phi = (0.75, 0.90, 0.99)$. Relative to coverage rates for $\phi = 0$ there is little change in the $I(0)$ model for either the Bayes or MN-frequentist prediction sets. In the local-level model the largest difference is 4%, which occurs with $\phi = 0.99$ and $r = 1.40$. Thus, even highly persistent stochastic volatility has only a small effect on coverage in these experiments.

A possible explanation for this robustness is that these results hold "on average" over the realizations of the volatility process, but that coverage may differ considerably over in-sample realizations of $\ln(\sigma_t^2)$. That is, one might conjecture that when $\ln(\sigma_T^2)$ is smaller than the unconditional volatility coverage exceeds 90%, but when $\ln(\sigma_T^2)$ is larger than the unconditional volatility coverage is less than 90%. Thus the other columns of the table investigate the sensitivity of coverage to the realized value of $\ln(\sigma_T^2)$. Columns 5-10 show coverage rates conditional on $\ln(\sigma_T)$ being one standard deviation above or below its unconditional value. While these conditional coverage rates differ from the unconditional rates as expected, the effects are relatively small. For example, the conditional coverage rates for the MN-frequentist sets differ from the nominal rates by 4% in the local-level model for short

Table 6: Coverage of 90% Prediction Sets for Models with Stochastic Volatility

a. $X_t \sim$ I(0)

| $r$ | | | unconditional | | | | $ln(\sigma_T) = ln(\sigma) + 1$sd | | | | $ln(\sigma_T) = ln(\sigma) - 1$sd | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\phi$ | | | | | | | | | | | | |
| | 0.0 | | 0.75 | 0.90 | 0.99 | | 0.75 | 0.90 | 0.99 | | 0.75 | 0.90 | 0.99 |
| | *Bayes Prediction Sets* | | | | | | | | | | | | |
| 0.05 | 0.89 | | 0.89 | 0.89 | 0.88 | | 0.88 | 0.86 | 0.84 | | 0.91 | 0.93 | 0.93 |
| 0.10 | 0.91 | | 0.91 | 0.90 | 0.90 | | 0.90 | 0.89 | 0.86 | | 0.91 | 0.93 | 0.94 |
| 0.60 | 0.94 | | 0.94 | 0.94 | 0.93 | | 0.94 | 0.94 | 0.93 | | 0.94 | 0.94 | 0.94 |
| 1.00 | 0.95 | | 0.95 | 0.94 | 0.94 | | 0.94 | 0.94 | 0.94 | | 0.95 | 0.95 | 0.94 |
| 1.40 | 0.95 | | 0.95 | 0.95 | 0.94 | | 0.95 | 0.95 | 0.94 | | 0.95 | 0.95 | 0.94 |
| | *MN-Frequentist Prediction Sets* | | | | | | | | | | | | |
| 0.05 | 0.91 | | 0.91 | 0.91 | 0.91 | | 0.90 | 0.89 | 0.87 | | 0.93 | 0.95 | 0.95 |
| 0.10 | 0.91 | | 0.91 | 0.91 | 0.91 | | 0.91 | 0.89 | 0.87 | | 0.92 | 0.93 | 0.95 |
| 0.60 | 0.96 | | 0.96 | 0.96 | 0.95 | | 0.96 | 0.95 | 0.95 | | 0.96 | 0.96 | 0.96 |
| 1.00 | 0.97 | | 0.97 | 0.97 | 0.96 | | 0.97 | 0.97 | 0.96 | | 0.97 | 0.97 | 0.97 |
| 1.40 | 0.97 | | 0.97 | 0.97 | 0.97 | | 0.97 | 0.97 | 0.97 | | 0.97 | 0.97 | 0.97 |

b. $X_t \sim$ Local Level Model

| $r$ | | | unconditional | | | | $ln(\sigma_T) = ln(\sigma) + 1$sd | | | | $ln(\sigma_T) = ln(\sigma) - 1$sd | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\phi$ | | | | | | | | | | | | |
| | 0.0 | | 0.75 | 0.90 | 0.99 | | 0.75 | 0.90 | 0.99 | | 0.75 | 0.90 | 0.99 |
| | *Bayes Prediction Sets* | | | | | | | | | | | | |
| 0.05 | 0.91 | | 0.91 | 0.90 | 0.89 | | 0.86 | 0.83 | 0.82 | | 0.95 | 0.97 | 0.96 |
| 0.10 | 0.90 | | 0.89 | 0.89 | 0.88 | | 0.85 | 0.82 | 0.81 | | 0.94 | 0.96 | 0.95 |
| 0.60 | 0.86 | | 0.86 | 0.85 | 0.84 | | 0.85 | 0.83 | 0.79 | | 0.87 | 0.87 | 0.89 |
| 1.00 | 0.86 | | 0.86 | 0.85 | 0.82 | | 0.85 | 0.83 | 0.80 | | 0.86 | 0.86 | 0.85 |
| 1.40 | 0.86 | | 0.86 | 0.85 | 0.82 | | 0.85 | 0.83 | 0.81 | | 0.86 | 0.85 | 0.83 |
| | *MN-Frequentist Prediction Sets* | | | | | | | | | | | | |
| 0.05 | 0.93 | | 0.93 | 0.92 | 0.91 | | 0.89 | 0.87 | 0.86 | | 0.97 | 0.98 | 0.97 |
| 0.10 | 0.91 | | 0.91 | 0.90 | 0.89 | | 0.87 | 0.84 | 0.83 | | 0.95 | 0.96 | 0.96 |
| 0.60 | 0.91 | | 0.91 | 0.90 | 0.89 | | 0.90 | 0.88 | 0.85 | | 0.91 | 0.91 | 0.92 |
| 1.00 | 0.92 | | 0.92 | 0.91 | 0.89 | | 0.91 | 0.90 | 0.87 | | 0.92 | 0.92 | 0.91 |
| 1.40 | 0.93 | | 0.93 | 0.92 | 0.90 | | 0.92 | 0.91 | 0.89 | | 0.93 | 0.92 | 0.91 |

Notes: The table shows the finite-sample coverage rates for Bayes and *MN*-Frequentist prediction sets for the *I*(0) model with stochastic volatility (panel a) and the local-level model (panel b). Parameter values for the models are discussed in the text.

run predictions ($r = 0.05$) and persistent stochastic volatility ($\phi = 0.99$), but by only 1% as the forecast horizon increases to $r = 1.4$.

## 7.2 Multivariate Prediction Sets

In a number of contexts it would be desirable to extend the analysis to a multivariate framework, where several time series are modelled and forecast jointly. One motivation would be that allowing for the richer information set from several variables might increase forecast accuracy. Another, perhaps even more compelling motivation is an interest in studying the joint uncertainty of long-run forecasts of several variables. For instance, one might want to forecast several components of GDP growth jointly (such as population growth, employment growth and productivity growth), and then use them for 'uncertainty accounting' purposes, that is to identify the major contributors to overall long-run GDP growth uncertainty.[15] Another application would be the joint uncertainty about long-run stock returns and consumption, which features prominently in asset pricing models that focus on long-run risk (Bansal and Yaron (2004) and Hansen, Heaton, and Li (2008)).

Conceptually, it is fairly straightforward to extend our framework to the multivariate setting. With $Y_T$, $\bar{x}_{1:T}$ and $X_{T,1:q}$ now $1 \times k$ and $q \times k$ matrices, with each column representing the transformations of Section 2 for one series, one would expect that under suitable conditions, the analogue of the central limit theorem presented in the appendix to hold, so that

$$T^{1-\alpha} \operatorname{vec} \begin{pmatrix} Y_T \\ X_{T,1:q} \end{pmatrix} \Rightarrow \mathcal{N}(0, \Sigma) \tag{15}$$

where the elements of $\Sigma$ are functions of some $k$ dimensional local-to-zero spectrum. Location and scale invariance then corresponds to the requirement that if $y \in A(\overline{x}_{1:T}, X_{T,1:q})$, then also $m + by \in A(m + \overline{x}_{1:T}b, X_{T,1:q}b)$, for any $m \in \mathbb{R}^k$ and full rank $k \times k$ matrix $b$. It is not hard to see that under this invariance, one can restrict attention to the problem of forecasting $Y_T^s = Y_T(X_{T,1:q}^k)^{-1}$ from $X_T^s = X_{T,1:q}(X_{T,1:q}^k)^{-1}$, where $X_{T,1:q}^k$ are the first $k$ rows of the $q \times k$ matrix $X_{T,1:q}$. Thus, with some parametric assumption on the shape of the spectral density, one could in principle proceed as in the univariate problem.

As a practical matter, though, there are two substantial difficulties. First, even with moderate values of $k$, the number of parameters describing the multivariate local-to-zero spectrum becomes fairly large. For instance, with $k = 3$, even the strong assumption that

[15] For example, Gordon (2003) provides a seven–way decomposition of the 75-year ahead forecast of GDP growth.

under some rotation, the three variables are independent with a local-to-zero spectrum in the *abd*-family, yields a 9-dimensional parameter space. The numerical determination of weighted expected volume minimizing prediction sets then becomes computationally challenging.

These numerical difficulties are compounded by the form of the density of $X^s$ and $Y^s$ that is induced by the normal distribution (15). In the scalar case considered in this paper these densities are in closed form. But in the multivariate setting, Theorem 1 of Müller and Watson (2011) yields an expression for the density that involves an integral of dimension $k^2$ unless $\Sigma$ has a specific Kronecker form. It is an interesting question for future research how to overcome these two computational challenges.

# 8 Summary and Concluding Remarks

This paper has considered the problem of quantifying uncertainty about long-run predictions using prediction sets that contain the realized future value of a variable of interest with prespecified probability. The long-run nature of the problem both simplifies and complicates the problem relative to short-run predictions. The problem is simplified because of our focus on forecasting long-run averages using a relatively small number of (low-frequency) weighted averages of the sample data. We show that these averages have an approximate joint normal distribution, which greatly simplifies the prediction problem. However, the prediction problem is complicated because the covariance matrix of the limiting normal distribution depends on the spectrum over very low frequencies, and there is limited sample information about its shape. Uncertainty about the low-frequency characteristics of the stochastic process is then an important component of the uncertainty about long-run predictions.

We proposed a flexible parametric model (the *bcd*-model) to characterize the shape of the spectrum at low frequencies. Uncertainty about the shape them becomes equivalent to uncertainty about the values of the *bcd*-parameters. Incorporating this parameter uncertainty into prediction uncertainty is straightforward in a Bayesian framework, and we provide the details in the context of the long-run prediction problem. However, because of the paucity of sample information about these long-run parameters, the resulting Bayes prediction sets may depend importantly on the specifics of the prior. This motivates us to consider frequentist prediction sets which, by construction, control coverage uniformly over all values of the parameters. We construct minimum expected volume frequentist prediction sets using an approximate "least favorable distribution" for the parameters, and we generalize these to conditionally sensible frequentist prediction sets using ideas from Müller and Norets (2012).

We apply these methods and construct prediction sets for nine macroeconomic time series for forecast horizons of up to 75 years. In general, we found that for many series, the prediction sets are wider than those that one obtains from the I(0) model, but not quite as wide as one would obtain from, say, the $I(1)$ model. From a statistical point of view, this underlines the importance of modelling the spectral shape at low frequencies in a flexible manner. Substantively, it demonstrates that even after accounting for a wide variety of potential long-run instabilities and dependencies, it is still possible to make informative statements about (very) long-run forecasts.

# References


Baillie, R. T. (1996): "Long Memory Processes and Frational Integration in Econometrics," *Journal of Econometrics*, 73, 5–59.

Bansal, R., and A. Yaron (2004): "Risks in the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Barro, R. J. (2006): "Rare disasters and asset markets in the twentieth century," *The Quarterly Journal of Economics*, 121(3), 823–866.

Beran, J. (1994): *Statistics for Long-Memory Processes*. Chapman and Hall, London.

Campbell, J. Y., and L. M. Viceira (1999): "Consumption and Portfolio Decisions When Expected Returns are Time Varying," *Quarterly Journal of Economics*, 114, 433–495.

Carter, S. B., S. S. Gartner, M. R. Haines, A. L. Olmstead, R. Sutch, and G. Wright (2006): *Historical Statistics of the United States, Millennial Edition Online*. Cambridge University Press.

Chan, N. H., and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

Clark, P. K. (1987): "The cyclical component of US economic activity," *The Quarterly Journal of Economics*, 102(4), 797–814.

Cogley, T., G. Primiceri, and T. J. Sargent (2009): "Inflation-gap persistence in the U.S.," *American Economic Journal: Macroeconomics*, 2, 43–69.

Congressional Budget Office (2005): "Quantifying Uncertainty in the Analyis of Long-Term Social Security Projections," *CBO Background paper*.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, New York.

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

Doornik, J. A., and M. Ooms (2004): "Inference and Forecasting for ARFIMA Models With an Application to US and UK Inflation," *Studies in Nonlinear Dynamics & Econometrics*, 8.

Elliott, G. (2006): "Forecasting with Trending Data," in *Handbook of Economic Forecasting, Volume 1*, ed. by C. W. J. G. G. Elliott, and A. Timmerman. North-Holland.

Elliott, G., U. K. Müller, and M. W. Watson (2013): "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Working Paper, Princeton University*.

Fernald, J. (2012): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," *FRBSF Working Paper 2012-9*.

Gordon, R. J. (1998): "Foundations of the Goldilocks economy: supply shocks and the time-varying NAIRU," *Brookings Papers on Economic Activity*, 29(2), 297–346.

Granger, C. W., and Y. Jeon (2007): "Long-term forecasting and evaluation," *International Journal of Forecasting*, 23(4), 539 – 551.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260–302.

Harvey, A. C. (1985): "Trends and cycles in macroeconomic time series," *Journal of Business & Economic Statistics*, 3(3), 216–227.

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

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

KEMP, G. C. R. (1999): "The Behavior of Forecast Errors from a Nearly Integrated AR(1) Model as Both Sample Size and Forecast Horizon Become Large," *Econometric Theory*, 15(2), pp. 238–256.

KIEFER, N., AND T. J. VOGELSANG (2002): "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory*, 18, 1350–1366.

——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164.

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

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

KITSUL, Y., AND J. H. WRIGHT (2012): "The Economics of Options-Implied Inflation Probability Density Functions," *Working Paper, Johns Hopkins University*.

LEE, R. (2011): "The Outlook for Population Growth," *Science*, 333, 569–573.

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. Springer, New York.

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

MÜLLER, U. K. (2012): "HAC Corrections for Strongly Autocorrelated Time Series," *Working paper, Princeton University*.

MÜLLER, U. K., AND A. NORETS (2012): "Credibility of Confidence Sets in Nonstandard Econometric Problems," *Working Paper, Princeton University*.

MÜLLER, U. K., AND M. W. WATSON (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016.

——— (2011): "Low-Frequency Robust Cointegration Testing," *Working paper, Princeton University*.

——— (2013): "The local-level local-to-unity fractional model," *Working Paper, Princeton University*.

Nelson, C. R., and G. W. Schwert (1977): "Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant," *The American Economic Review*, 67(3), 478–486.

Pastor, L., and R. F. Stambaugh (2012): "Are Stocks Really Less Volatile in the Long Run?," *Journal of Finance*, LXVII, 431–477.

Perron, P., and Z. Qu (2010): "Long-memory and level shifts in the volatility of stock market return indices," *Journal of Business & Economic Statistics*, 28(2), 275–290.

Pesavento, E., and B. Rossi (2006): "Small-Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons," *Journal of Applied Econometrics*, 21(8), 1135–1155.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (1998): "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs," *Journal of Econometrics*, 83, 21–56.

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

Raftery, A. E., N. Li, H. Sevcíkova, P. Gerland, and G. K. Heilig (2012): "Bayesian probabilistic population projections for all countries," *Proceedings of the National Academy of Sciences*.

Rietz, T. A. (1988): "The equity risk premium a solution," *Journal of monetary Economics*, 22(1), 117–131.

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

Siegel, J. (2007): *Stocks for the Long Run: The Definitive Guide to Financial Markey Returns and Long Term Investment Strategies*. McGraw-Hill.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H. (1991): "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459.

——— (1996): "VAR, Error Correction and Pretest Forecasts at Long Horizons," *Oxfor*, 58, 685–701.

——— (1997): "Cointegration, Long-Run Comovements, and Long-Horizon Forecasting," in *Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society*, ed. by D. Kreps, and K. Wallis, pp. 34–60, Cambridge. Cambridge University Press.

STOCK, J. H., AND M. W. WATSON (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.

——— (2007): "Why Has Inflation Become Harder to Forecast?," *Journal of Money, Credit, and Banking*, 39, 3–33.

——— (2010): "Modeling Inflation After the Crisis," *Federal Reserve Bank of Kansas City Jackson Hole Symposium*.

SUMMERS, L. H. (1986): "Does the stock market rationally reflect fundamental values?," *The Journal of Finance*, 41(3), 591–601.

WATSON, M. W. (1986): "Univariate detrending methods with stochastic trends," *Journal of monetary economics*, 18(1), 49–75.

# 9 Appendix

## 9.1 A Guide to Computing Long-Run Prediction Sets

This appendix section summarizes the key steps for computing the long-run prediction sets. To review notation, the sample data are $\{x_t\}_{t=1}^T$, $\overline{x}_{T+1:T+h} = h^{-1}\sum_{t=1}^h x_{T+t}$ is the variable to be forecast, and $\theta = (a, b, d)$ denotes a vector of parameters that characterize the shape of the low-frequency spectrum. To simplify the discussion, discretize the parameter space so that $\theta$ takes on the values $\theta_i$, $i = 1, ..., n_\theta$. Let $\overline{x}_{1:T}$ denote the in-sample mean of $x_t$, and (from equation (3)) let $X_{T,1:q} = (X_T(1), \ldots, X_T(q))'$ denote the first $q$ cosine-weighted averages of the sample data, $Y_T = \overline{x}_{T+1:T+h} - \overline{x}_{1:T}$ denote the difference between the out-of-sample and in-sample mean, and $X^s_{T,1:q} = X_{T,1:q}/\sqrt{X'_{T,1:q}X_{T,1:q}}$and $Y^s_T = Y_T/\sqrt{X'_{T,1:q}X_{T,1:q}}$ denote scaled versions of $X_{T,1:q}$ and $Y_T$. A level $1-\alpha$ prediction set $A(X_T)$ satisfies $P(\overline{x}_{T+1:T+h} \in A(X_T)) \geq 1-\alpha$, at least asymptotically, based on the approximation that under $\theta_i$, $T^{1-\alpha}(X'_{T,1:q}, Y_T)' \Rightarrow (X'_{1:q}, Y)' \sim \mathcal{N}(0, \Sigma(\theta_i))$.

The prediction sets are computed as follows[16]:

1. Compute the sample averages $\overline{x}_{1:T}$, $X_{T,1:q}$ (from (3)), and $X^s_{T,1:q} = X_{T,1:q}/\sqrt{X'_{T,1:q}X_{T,1:q}}$, where $q$ is chosen to capture the low-frequency variability in $x_t$. (In our empirical work we set $q = 12$, which captures variability for periods greater than 10 years in post-WWII samples.)

2. For each $\theta_i$ compute the covariance matrix $\Sigma(\theta_i)$. From equation (6) the $j,k$'th element of $\Sigma(\theta_i)$ is $\Sigma(\theta_i)_{jk} = 2\int_0^\infty S_i(\omega)w_{jk}(\omega)d\omega$, where $S_i(\omega) = (\omega^2 + c_i^2)^{-d_i} + b_i^2$ is the low-frequency (pseudo) spectrum for $\theta_i = (b_i, c_i, d_i)$ and $w_{jk}(\omega)$ are weights given in the paragraph below equation (6).

3. To compute the Bayes prediction set with coverage $1-\alpha$:

   (a) Specify a prior for $\theta$. In Section 4, the prior distribution is denoted $\Gamma$, and thus let $\gamma_i$ denote the prior probability that $\theta = \theta_i$.

   (b) Compute the (asymptotic) marginal likelihood of $X^s_{T,1:q}$, which from equation (8) is $f^\Gamma_{X^s}(X^s_{T,1:q}) = \frac{1}{2}\Gamma(\frac{q}{2})\pi^{-q/2}\sum_{i=1}^{n_\theta}\gamma_i|\Sigma_{X,i}|^{-1/2}(X^{s\prime}_{T,1:q}\Sigma^{-1}_{X,i}X^s_{T,1:q})^{-q/2}$, where $\Sigma_{X,i}$ is the upper $q \times q$ submatrix of $\Sigma(\theta_i)$ computed in step 2.

   (c) Compute a prediction set for $Y^s_T$.[17]

[16] Matlab programs to carry out each of these steps are available in the replication files for this paper.

[17] Steps (3.c.ii) and (3.c.iii) can be carried out using a discrete ("histogram") approximation for the predictive distribution using the density from (3.c.i).

i. The (asymptotic) predictive density for $Y_T^s$ is $f^{\Gamma}_{Y^s|X^s}(y_T^s) = f^{\Gamma}_{(X^s,Y^s)}(X^s_{T,1:q}, y^s_T)/f^{\Gamma}_{X^s}(X^s_{T,1:q})$, where $f^{\Gamma}_{(X^s,Y^s)}(X^s_{T,1:q}, y^s_T) = \frac{1}{2}\Gamma(\frac{q+1}{2})\pi^{-(q+1)/2}\sum_{i=1}^{n_\theta}\gamma_i|\Sigma(\theta_i)|^{-1/2}(W_T^{s\prime}\Sigma(\theta_i)^{-1}W_T^s)^{-(q+1)/2}$ with $W_T^s = (X^{s\prime}_{T,1:q}, y^s_T)'$.

ii. The highest predictive density (HPD) prediction set for $Y_T^s$ is $A^{Y^s}_{Bayes,HPD} = \{y^s | f^{\Gamma}_{Y^s|X^s}(y^s) > \text{cv}\}$, where cv is chosen so that $\int_{A^{Y^s}_{Bayes,HPD}} f^{\Gamma}_{Y^s|X^s}(y^s)dy^s = 1-\alpha$.

iii. The equal-tailed (ET) prediction set for $Y_T^s$ is $A^{Y^s}_{Bayes,ET} = (q^{\Gamma}_{\alpha/2}, q^{\Gamma}_{1-\alpha/2})$, where $q^{\Gamma}_{\delta}$ is the $\delta$'th quantile of the predictive distribution with density $f^{\Gamma}_{Y^s|X^s}(y^s)$.

(d) Compute the prediction sets for $\overline{x}_{T+1:T+h}$

i. Highest predictive density prediction set: $\{y | y = \overline{x}_{1:T} + y^s\sqrt{X'_{T,1:q}X_{T,1:q}}$ for $y^s \in A^{Y^s}_{Bayes,HPD}\}$.

ii. Equal-tailed prediction set: $\{y | y = \overline{x}_{1:T} + y^s\sqrt{X'_{T,1:q}X_{T,1:q}}$ for $y^s \in A^{Y^s}_{Bayes,ET}\}$.

4. To compute the MN-frequentist prediction set with coverage $1-\alpha$:

(a) Specify the weighting function for $\theta$ for the weighted average volume of the prediction set. The distribution is denoted $\Gamma$ in section 4, and thus let $\gamma_i$ denote the weight placed on $\theta = \theta_i$.

(b) Compute the Bayes prediction set for $\overline{x}_{T+1:T+h}$ using $\Gamma$ as the prior, as in step 3.d. Denote this set $A^{Bayes}_{\Gamma}$ (which can be either of the $HPD$ or $ET$ Bayes sets).

(c) Compute the least favorable distribution $\Lambda$ and critical value $\text{cv}_\Lambda$.[18] Let $\lambda_i$ denote the weight placed on $\theta = \theta_i$.

(d) Compute the set $A^{Y^s}_{\Gamma,\Lambda} = \{y^s | f^{\Lambda}_{(X^s,Y^s)}(X^s_{T,1:q}, y^s)/f^{\Gamma}_{X^s}(X^s_{T,1:q}) > \text{cv}_\Lambda\}$, where $f^{\Lambda}_{(Y^s,X^s)}(X^s_{T,1:q}, y^s)$ is given in step (3.c.i) with $\lambda_i$ replacing $\gamma_i$.

(e) Compute the set $A^Y_{\Gamma,\Lambda} = \{y | y = \overline{x}_{1:T} + y^s\sqrt{X'_{T,1:q}X_{T,1:q}}$ for $y^s \in A^{Y^s}_{\Gamma,\Lambda}\}$.

(f) The MN-frequentist prediction set is given by $A^{MN}_{\Gamma} = A^Y_{\Gamma} \cup A^{Bayes}_{\Gamma}$.

## 9.2 Central Limit Theorem Used in Section 3

**Theorem 1** *Let* $\Delta x_{T,t} = \sum_{s=-\infty}^{\infty} c_{T,s}\varepsilon_{t-s}$. *Suppose that*

*(i)* $\{\varepsilon_t, \mathcal{F}_t\}$ *is a martingale difference sequence with* $E(\varepsilon_t^2) = 1$, $\sup_t E(|\varepsilon_t|^{2+\delta}) < \infty$ *for some* $\delta > 0$, *and*

$$E(\varepsilon_t^2 - 1|\mathcal{F}_{t-m}) \leq \xi_m \tag{16}$$

[18] This requires the numerical methods discussed in Appendix 9.4.

*for some sequence* $\xi_m \to 0$.

*(ii) For every* $\epsilon > 0$ *the exists an integer* $L_\epsilon > 0$ *such that* $\limsup_{T\to\infty} T^{-1} \sum_{l=L_\epsilon T+1}^{\infty} \left(T \sup_{|s|\geq l} |c_{T,s}|\right)^2 < \epsilon$.

*(iii)* $\sum_{s=-\infty}^{\infty} c_{T,s}^2 < \infty$ *(but not necessarily uniformly in* $T$*). The spectral density of* $\Delta x_{T,t}$ *thus exists; denote it by* $F_T : [-\pi, \pi] \mapsto \mathbb{R}$.

*(iii.a) Assume that there exists a function* $S : \mathbb{R} \mapsto \mathbb{R}$ *such that* $\omega \mapsto \omega^2 S(\omega)$ *is integrable, and for all fixed* $K$,

$$\int_0^K |F_T(\frac{\omega}{T}) - \omega^2 S(\omega)| d\omega \to 0. \tag{17}$$

*(iii.b) For every diverging sequence* $K_T \to \infty$

$$T^{-3} \int_{K_T/T}^{\pi} F_T(\lambda) \lambda^{-4} d\lambda = \int_{K_T}^{\pi T} F_T(\omega/T) \omega^{-4} d\omega \to 0. \tag{18}$$

*(iii.c)*

$$T^{-3/2} \int_{1/T}^{\pi} F_T(\lambda)^{1/2} \lambda^{-2} d\lambda = T^{-1/2} \int_1^{\pi T} F_T(\omega/T)^{1/2} \omega^{-2} d\omega \to 0. \tag{19}$$

*(iv) For some fixed integer* $H$*, the bounded and integrable function* $g : [0, H] \mapsto \mathbb{R}$ *satisfies* $\int_0^H g(s) ds = 0$*, and with* $G(r) = \int_0^r g(s) ds$*, for some constant* $C$, $|\sum_{t=1}^{HT} e^{-\mathrm{i}\lambda t} G(\frac{t-1}{T})| \leq C\lambda^{-2} T^{-1}$ *uniformly in* $T$ *and* $\lambda$.

*Then*

$$T^{-1/2} \int_0^H g(s) x_{T,[sT]+1} ds \Rightarrow \mathcal{N}(0, \int_{-\infty}^{\infty} S(\omega) \left| \int_0^H e^{-\mathrm{i}\omega s} g(s) ds \right|^2 d\omega) \tag{20}$$

*where* $x_{T,t} = \sum_{s=1}^{t} \Delta x_{T,s}$.

**Remarks:** Note that the linear process $\Delta x_{T,t}$ is not restricted to be causal. The m.d.s. structure of the driving errors $\varepsilon_t$ in assumption (i) allows for some departures from strict stationarity. It also accommodates conditional heteroskedasticity, with the second order dependence limited by the mixingale condition (16).

With the pseudo-spectrum of $x_{T,t}$ defined as $R_T(\lambda) = F_T(\lambda)/|1 - e^{-\mathrm{i}\lambda}|^2$, assumption (iii.a) is equivalent to

$$T^2 \int_0^K \omega^2 |T^{-2} R_T(\frac{\omega}{T}) - S(\omega)| d\omega \to 0 \tag{21}$$

since for any fixed $K$, $\sup_{0\leq\omega\leq K} |T^{-2} \frac{\omega^2}{|1-e^{-\mathrm{i}\omega/T}|^2} - 1| \to 0$. Thus, (iii.a) is equivalent to the convergence of the pseudo-spectrum of $\{x_{T,t}\}$ to $S$ in a $T^{-1}$ neighborhood of the origin in the sense of (21).

To better understand the role of assumptions (ii) and (iii), consider some leading examples. Suppose first that $\Delta x_{T,t}$ is causal and weakly dependent with exponentially decaying $c_{T,s}$, $|c_{T,s}| \leq C_0 e^{-C_1 s}$ for some $C_0, C_1 > 0$, as would arise in causal and invertible ARMA models of any fixed

and finite order. Then $T^{-1}\sum_{l=LT+1}^{\infty}\left(T\sup_{|s|\geq l}|c_{T,s}|\right)^2 \to 0$ for any $L>0$, $\omega^2 S(\omega)$ is constant and equal to the long-run variance of $\Delta x_{T,t}$, and (18) and (19) hold, since $F_T$ is bounded, $\int_{K_T}^{\infty}\omega^{-4}d\omega \to 0$ for any $K_T \to \infty$ and $\int_1^{\infty}\omega^{-2}d\omega < \infty$.

Second, suppose $\Delta x_{T,t}$ is fractionally integrated with parameter $d \in (-1/2, 1/2)$ (corresponding to $x_{T,t}$ being fractionally integrated of order $d+1$). With $\Delta x_{T,t}$ scaled by $T^{-d}$, $c_{T,s} \approx C_0 T^{-d} s^{d-1}$, so that $T^{-1}\sum_{l=LT+1}^{\infty}\left(T\sup_{|s|\geq l}|c_{T,s}|\right)^2 \to \int_L^{\infty} s^{2d-2}ds$, which can be made arbitrarily small by choosing $L$ large. Further, for $\lambda$ close to zero, $F_T(\lambda) \approx C_0^2(\lambda T)^{-2d}$, so that $\omega^2 S(\omega) = C_0^2\omega^{-2d}$, and (18) and (19) are seen to hold under weak assumptions about higher frequency properties of $\Delta x_{T,t}$. For instance, even integrable poles in $F_T$ at frequencies other than zero can be accommodated.

Third, suppose $x_{T,t}$ is an AR(1) process with local-to-unity coefficient $\rho_T = 1 - c/T$ and unit innovation variance. Then $c_{T,0} = 1$ and $c_{T,s} = -(1-\rho_T)\rho_T^s$, $s > 0$. Thus $T^{-1}\sum_{l=LT+1}^{\infty}\left(T\sup_{|s|\geq l}|c_{T,s}|\right)^2 \to c^2\int_L^{\infty} e^{-2cs}ds$, which can be made arbitrarily small by choosing $L$ large. Further, $F_T(\lambda) = |1-e^{-\mathrm{i}\lambda}|^2/|1-\rho_T e^{-\mathrm{i}\lambda}|^2$, which is seen to satisfy (17) with $S(\omega) = 1/(\omega^2+c^2)$. Conditions (18) and (19) also hold in this example, since $F_T(\lambda) \leq 1$.

As a final example, suppose $\Delta x_{T,t} = T\varepsilon_t - T\varepsilon_{t-1}$ (inducing $x_{T,t}$ to be i.i.d. conditional on $\varepsilon_0$). Here $F_T(\lambda) = T^2|1-e^{-\mathrm{i}\lambda}|^2 = 4T^2\sin(\lambda/2)^2$, so that $S(\omega) = 1$, and (18) evaluates to $4\int_{K_T}^{\pi T} T^2\sin(\omega/2T)^2\omega^{-4}d\omega \leq \int_{K_T}^{\pi T}\omega^{-2}d\omega \to 0$, and (18) to $2T^{-1/2}\int_1^{\pi T} T\sin(\frac{1}{2}\omega/T)\omega^{-2}d\omega \leq T^{-1/2}\int_1^{\pi T}\omega^{-1}d\omega \to 0$, where the inequalities follow from $\sin(\lambda) \leq \lambda$ for all $\lambda \geq 0$.

Assumption (iv) of Theorem 1 specifies the sense in which the weights $g$ in (20) are smooth, so that the properties of the $g(s)$-weighted average of $x_{T,[sT]+1}$ are wholly determined by the low-frequency properties of $x_{T,t}$. The number $H$ is assumed to be an integer to ease notation. Note that a constant $g$ would not satisfy assumption (iv), as it does not integrate to zero, but all functions of interest in the context of this paper do.

**Lemma 1** *For some $0 < r < H-1$, let $g_{q+1} : [0,H] \mapsto \mathbb{R}$ equal $g_{q+1}(s) = -\mathbf{1}[0 \leq s \leq 1] + r^{-1}\mathbf{1}[1 < s \leq 1+r]$ and let $g_j : [0,H] \mapsto \mathbb{R}$ equal to $g_j(s) = \mathbf{1}[s \leq 1]\sqrt{2}\cos(\pi j s)$ for $j = 1,\ldots,q$. Then for any fixed and real $\lambda_j$, $g(s) = \sum_{j=1}^{q+1}\lambda_j g_j(s)$ satisfies assumption (iv) of Theorem 1.*

**Proof.** It is clearly sufficient to show that each $g_j$ satisfies the assumption. This follows from a direct computation. ■

The implication of Theorem 1 that is of interest for Section 3 follows from the following Corollary.

**Corollary 1** *Suppose $g_1,\ldots,g_{q+1}$ are as in Lemma 1. Then under the assumptions of Theorem 1*

*(i)-(iii),*

$$T^{-1/2}\int_0^H \begin{bmatrix} g_1(s) \\ \vdots \\ g_q(s) \\ g_{q+1}(s) \end{bmatrix} x_{T,\lfloor sT\rfloor+1} ds \Rightarrow \mathcal{N}(0,\Sigma)$$

*where* $\Sigma_{j,k} = \int_{-\infty}^{\infty} S(\omega)\left(\int_0^H e^{-\mathrm{i}\omega s} g_j(s)ds\right)\left(\int_0^H e^{\mathrm{i}\omega s} g_k(s)ds\right)d\omega$ *for* $j,k = 1,\ldots,q+1$.

**Proof.** Follows from Theorem 1, Lemma 1 and the Cramer-Wold device via

$$\left|\int_0^H e^{-\mathrm{i}\omega s}\left(\sum_{j=1}^{q+1}\lambda_j g_j(s)\right)ds\right|^2 = \sum_{j,k=1}^{q+1}\lambda_j\lambda_k\left(\int_0^H e^{-\mathrm{i}\omega s} g_j(s)ds\right)\left(\int_0^H e^{\mathrm{i}\omega s} g_k(s)ds\right).$$

■

We now introduce some notation that is used in the proof of Theorem 1, and in auxiliary Lemmas. We first state these Lemmas and then the proof of Theorem 1.

**Notation:** Define $\sigma_T^2 = \mathrm{Var}[T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds]$. Note that with $\tilde{g}_{T,t} = T\int_{(t-1)/T}^{t/T} g(s)ds$ and $\tilde{G}_{T,t} = T^{-1}\sum_{s=1}^{t-1}\tilde{g}_{T,t} = \sum_{s=1}^{t-1}\int_{(s-1)/T}^{s/T} g(s)ds = \int_0^{(t-1)/T} g(s)ds = G(\frac{t-1}{T})$, we find using summation by parts

$$\begin{aligned} T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds &= T^{-3/2}\sum_{t=1}^{HT}\tilde{g}_{T,t}x_{T,t} \\ &= T^{-1/2}\tilde{G}_{T,HT+1}x_{T,HT} - T^{-1/2}\sum_{t=1}^{HT}\tilde{G}_T(t/T)\Delta x_{T,t} \\ &= -T^{1/2}\sum_{t=1}^{HT} G(\frac{t-1}{T})\Delta x_{T,t} \end{aligned}$$

since $\tilde{G}_{T,HT+1} = G(H) = 0$.

**Lemma 2** $\sigma_T^2 \to \sigma^2 = \int_{-\infty}^{\infty} S(\omega)\left|\int_0^H e^{-\mathrm{i}\omega s}g(s)ds\right|^2 d\omega$.

**Proof.** Let $\gamma_T$ be the autocovariances of $\Delta x_{T,t}$. We find

$$\begin{aligned} \mathrm{Var}[T^{-1/2}\sum_{t=1}^{HT} G(\frac{t-1}{T})\Delta x_{T,t}] &= T^{-1}\sum_{j,k=1}^{HT}\gamma_T(k-j)G(\frac{k-1}{T})G(\frac{j-1}{T}) \\ &= T^{-1}\sum_{j,k=1}^{HT}\left(\int_{-\pi}^{\pi} e^{\mathrm{i}\lambda(k-j)}F_T(\lambda)d\lambda\right)G(\frac{k-1}{T})G(\frac{j-1}{T}) \end{aligned}$$

$$= \quad T^{-1}\int_{-\pi}^{\pi} F_T(\lambda)\left|\sum_{t=1}^{HT} e^{\mathrm{i}\lambda t}G(\frac{t-1}{T})\right|^2 d\lambda.$$

Now for any fixed $K$, we will show

$$\begin{aligned} T^{-1}\int_{-K/T}^{K/T} F_T(\lambda)\left|\sum_{t=1}^{HT} e^{\mathrm{i}\lambda t}G(\frac{t-1}{T})\right|^2 d\lambda &= \int_{-K}^{K} F_T(\frac{\omega}{T})\left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T}G(\frac{t-1}{T})\right|^2 d\omega \\ &\to \int_{-K}^{K} S(\omega)\omega^2\left|\int_0^H e^{\mathrm{i}\omega s}G(s)ds\right|^2 d\omega. \end{aligned} \tag{22}$$

First note that for any two complex numbers $a, b$, $|a-b| \geq ||a|-|b||$, so that $||a|^2-|b|^2| = |(|a|+|b|)(|a|-|b|)| \leq (|a|+|b|)|a-b|$. Thus,

$$\left|\left|\int_0^H e^{\mathrm{i}\omega s}G(s)ds\right|^2 - \left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T}G(\frac{t-1}{T})\right|^2\right| \leq$$

$$2\sup_s|G(s)|\cdot\left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T}G(\frac{t-1}{T}) - \int_0^H e^{\mathrm{i}\omega s}G(s)ds\right|$$

and

$$\begin{aligned} &\left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T}G(\frac{t-1}{T}) - \int_0^H e^{\mathrm{i}\omega s}G(s)ds\right| \\ \leq\ &\int_0^H |e^{\mathrm{i}\omega s}G(s) - e^{\mathrm{i}\omega[sT]/T}G([sT]/T)|ds \\ =\ &\int_0^H |e^{\mathrm{i}\omega s}(G(s)-G([sT]/T)) + G([sT]/T)(e^{\mathrm{i}\omega s}-e^{\mathrm{i}\omega[sT]/T})|ds \\ \leq\ &\int_0^H |G(s)-G([sT]/T)|ds + \sup_s|G(s)|\int_0^H |1-e^{\mathrm{i}\omega([sT]/T-s)}|ds. \end{aligned}$$

Since for any real $a$, $|1-e^{\mathrm{i}a}| < |a|$, $\sup_{|\omega|\leq K}|1-e^{\mathrm{i}\omega(\lfloor sT\rfloor/T-s)}| \leq K/T$, and also $|G(s)-G([sT]/T)| \leq T^{-1}\sup_s|g(s)|$. Thus,

$$\sup_{|\omega|\leq K}\left|\left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T}G(\frac{t-1}{T})\right|^2 - \left|\int_0^H e^{\mathrm{i}\omega s}G(s)ds\right|^2\right| \to 0$$

and (22) follows from assumption (iii.a) by straightforward arguments.

Further, since $\int_0^H e^{\mathrm{i}\omega s}g(s)ds = -\mathrm{i}\omega\int_0^H e^{\mathrm{i}\omega s}G(s)ds$,

$$\int_{-K}^{K} S(\omega)\omega^2\left|\int_0^H e^{\mathrm{i}\omega s}G(s)ds\right|^2 d\omega = \int_{-K}^{K} S(\omega)\left|\int_0^H e^{\mathrm{i}\omega s}g(s)ds\right|^2 d\omega.$$

Thus, for any fixed $K$,

$$\delta_K(T) = T^{-1}\int_{-K/T}^{K/T} F_T(\lambda)\left|\sum_{t=1}^{HT} e^{\mathrm{i}\lambda t}G(\frac{t-1}{T})\right|^2 d\lambda - \int_{-K}^{K} S(\omega)\left|\int_0^H e^{\mathrm{i}\omega s}g(s)ds\right|^2 d\omega \to 0.$$

Now for each $T$, define $K_T$ as the largest integer $K \leq T$ for which $\sup_{T'\geq T}|\delta_K(T')| \leq 1/K$ (and zero if no such $K$ exists). Note that $\delta_K(T) \to 0$ for all fixed $K$ implies that $K_T \to \infty$, and by construction, also $\delta_{K_T}(T) \to 0$. The result now follows from

$$T^{-1}\int_{K_T/T}^{\pi} F_T(\lambda)\left|\sum_{t=1}^{HT} e^{\mathrm{i}\lambda t}G(\frac{t-1}{T})\right|^2 d\lambda \leq C^2T^{-3}\int_{K_T/T}^{\pi} F_T(\lambda)\frac{1}{\lambda^4}d\lambda \to 0$$

by assumptions (iii.b) and (iv). ■

**Lemma 3** $T^{-1/2}\sup_{t,T}|\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}| \to 0$.

**Proof.** Recall that for any two real, square integrable sequences $\{a_j\}_{j=-\infty}^{\infty}$ and $\{b_j\}_{j=-\infty}^{\infty}$, $\sum_{j=-\infty}^{\infty} a_jb_j = \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{A}(\lambda)\hat{B}^*(\lambda)d\lambda$, where $\hat{A}(\lambda) = \sum_{j=-\infty}^{\infty} a_je^{-\mathrm{i}\lambda j}$ and $\hat{B}^*(\lambda) = \sum_{j=-\infty}^{\infty} b_je^{\mathrm{i}\lambda j}$. Thus

$$\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t} = \sum_{j=1-t}^{HT-t} G(\frac{t+j-1}{T})c_{T,j} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\mathrm{i}\lambda t}\hat{G}_T(\lambda)\hat{C}_T^*(\lambda)d\lambda$$

where $\hat{C}_T(\lambda) = \sum_{j=-\infty}^{\infty} c_{T,j}e^{-\mathrm{i}\lambda j}$ and $\hat{G}_T(\lambda) = \sum_{j=1}^{HT} e^{-\mathrm{i}\lambda j}G(\frac{j-1}{T})$, so that $e^{\mathrm{i}\lambda t}\hat{G}_T(\lambda) = \sum_{j=1}^{HT} e^{-\mathrm{i}\lambda(j-t)}G(\frac{t-1}{T}) = \sum_{j=1-t}^{HT-t} e^{-\mathrm{i}\lambda j}G(\frac{t+j-1}{T})$, and $\hat{C}_T^*$ is the complex conjugate of $\hat{C}_T$. Now since $F_T(\lambda) = \frac{1}{2\pi}|\hat{C}_T(\lambda)|^2$, we find by the Cauchy-Schwarz inequality that

$$2\pi|\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}| = |\int_{-\pi}^{\pi} e^{\mathrm{i}\lambda t}\hat{G}_T(\lambda)\hat{C}_T^*(\lambda)d\lambda| \leq \sqrt{2\pi}\int_{-\pi}^{\pi}|\hat{G}_T(\lambda)|F_T(\lambda)^{1/2}d\lambda.$$

Also, since $|\hat{G}_T(\lambda)| \leq \sum_{j=1}^{HT}|G(\frac{j-1}{T})|$, we have

$$\begin{aligned}\int_0^{1/T}|\hat{G}_T(\lambda)|F_T(\lambda)^{1/2}d\lambda &\leq T^{-1}\sum_{j=1}^{HT}|G(\frac{j-1}{T})|\cdot\int_0^1 F_T(\omega/T)^{1/2}d\omega \\ &\to \int_0^H|G(s)|ds\cdot\int_0^1\omega S(\omega)^{1/2}d\omega < \infty\end{aligned}$$

where the convergence follows from assumption (iii.a). Furthermore, by assumption (iv),

$$\int_{1/T}^{\pi}|\hat{G}_T(\lambda)|F_T(\lambda)^{1/2}d\lambda \leq CT^{-1}\int_{1/T}^{\pi}F_T(\lambda)^{1/2}\lambda^{-2}d\lambda.$$

The result follows by assumption (iii.c). ■

**Lemma 4** *For every $\epsilon > 0$ there exists a $M > 0$ such that*

$$\mathrm{Var}[T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds + T^{-1/2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t] < \epsilon.$$

*For this $M$, $\sigma^2_{T,M} = \mathrm{Var}[T^{-1/2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t]$ satisfies $\limsup_{T\to\infty}|\sigma^2_{M,T} - \sigma^2| < \epsilon$.*

**Proof.** We have

$$\begin{aligned}
-\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds &= \sum_{t=1}^{HT}G(\frac{t-1}{T})\Delta x_{T,t} \\
&= \sum_{t=1}^{HT}G(\frac{t-1}{T})\sum_{s=-\infty}^{\infty}c_{T,s}\varepsilon_{t-s} \\
&= \sum_{t=-\infty}^{\infty}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t
\end{aligned}$$

so that, with $\bar{G} = \sup_{0\le s<H}|G(s)|$

$$\begin{aligned}
&\mathrm{Var}[T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds + T^{-1/2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t] \\
= \; & T^{-1}\sum_{l=-\infty}^{-MT-1}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-l}\right)^2 + T^{-1}\sum_{l=MT+1}^{\infty}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-l}\right)^2 \\
\le \; & \bar{G}^2T^{-1}\sum_{l=MT+1}^{\infty}\left(\sum_{j=1}^{HT}(|c_{T,j-l}| + |c_{T,j+l}|\right)^2 \\
\le \; & 4H^2\bar{G}^2T^{-1}\sum_{l=MT+1}^{\infty}\left(T\sup_{|s|\ge l-HT}|c_{T,s}|\right)^2 \\
= \; & 4H^2\bar{G}^2T^{-1}\sum_{l=(M-H)T+1}^{\infty}\left(T\sup_{|s|\ge l}|c_{T,s}|\right)^2
\end{aligned}$$

which can be made arbitrarily small by choosing $M$ large enough via assumption (ii).

The second claim follows directly from Lemma 2. ■

**Lemma 5** *For any large enough integer $M > 0$, $\sigma^{-1}_{M,T}T^{-1/2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t \Rightarrow \mathcal{N}(0,1)$.*

**Proof.** By the second claim in Lemma 4 and Lemma 2, $\sigma_{M,T} = O(1)$ and $\sigma_{M,T}^{-1} = O(1)$. By Theorem 24.3 in Davidson (1994), it thus suffices to show (a) $T^{-1/2}\sup_{1\leq t\leq HT}|\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\varepsilon_t| \xrightarrow{p} 0$ and (b) $T^{-1}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right)^2(\varepsilon_t^2 - 1) \xrightarrow{p} 0$.

(a) is implied by the Lyapunov condition via Davidson's (1994) Theorems 23.16 and 23.11. Thus, it suffices to show that

$$\sum_{t=-MT}^{MT} E[|T^{-1/2}\left(\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t|^{2+\delta}] \to 0.$$

Now

$$\begin{aligned}
&\sum_{t=-MT}^{MT} E\left[\left|T^{-1/2}\left(\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t\right|^{2+\delta}\right] \\
\leq\ & T^{-1-\delta/2}\sum_{t=-MT}^{MT}\left|\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right|^{2+\delta} E(|\varepsilon_t|^{2+\delta}) \\
\leq\ & (\sup_t E(|\varepsilon_t|^{2+\delta}))\cdot T^{-\delta/2}\sup_t\left|\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right|^{\delta}\cdot T^{-1}\sum_{t=-MT}^{MT}\left|\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right|^2 \\
=\ & (\sup_t E(|\varepsilon_t|^{2+\delta}))\cdot\left(T^{-1/2}\sup_t\left|\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right|\right)^{\delta}\cdot\sigma_{M,T}^2 \to 0
\end{aligned}$$

where the convergence follows from Lemma 3.

For (b), we apply Theorem 19.11 of Davidson (1994) with Davidson's $X$ and $c$ chosen as $X_{T,t}^D = c_{T,t}^D(\varepsilon_t^2 - 1)$ and $c_{T,t}^D = T^{-1}\left(\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right)^2$. Then $X_{T,t}^D/c_{T,t}^D = \varepsilon_t^2 - 1$ is uniformly integrable, since $\sup_t E(|\varepsilon_t|^{2+\delta}) < \infty$. Further,

$$\sum_{t=-MT}^{MT} c_{T,t}^D = T^{-1}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right)^2 = \sigma_{M,T}^2 = O(1)$$

and

$$\begin{aligned}
\sum_{t=-MT}^{MT}(c_{T,t}^D)^2 &= T^{-2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right)^4 \\
&= \sup_{1\leq t\leq HT}\left|T^{-1/2}\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right|^2\cdot T^{-1}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT} G(\frac{j-1}{T})c_{T,j-t}\right)^2
\end{aligned}$$

$$= \sup_{1\leq t\leq HT}\left|T^{-1/2}\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right|^2\cdot\sigma_{M,T}^2\to 0$$

where the convergence follows from Lemma 3. ■

**Proof of Theorem 1:**

By Lemmas 2, 4 and 5, for large enough $M$,

$$\sigma^{-1}T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds=\frac{\sigma_{M,T}}{\sigma}A_T+\frac{\sigma_{M,T}}{\sigma}B_T$$

where $A_T\Rightarrow\mathcal{N}(0,1)$, $\limsup_{T\to\infty}E(B_T^2)<\epsilon$ and $\limsup_{T\to\infty}|\sigma_{M,T}^2-\sigma^2|<\epsilon$. Thus, by Slutzky's Theorem, as $\epsilon\to 0$, the desired convergence in distribution follows. But $\epsilon$ was arbitrary, which proves the Theorem.

## 9.3 Density of Maximal Invariant of Section 3

Let $W=(X',Y)'$ and $U=\sqrt{X'X}$. Write $\mu_l$ for Lebesgue measure on $\mathbb{R}^l$, and $\nu_q$ for the surface measure of a $q$ dimensional unit sphere. For $x\in\mathbb{R}^q$, let $x=x^su$, where $x^s$ is a point on the surface of a $q$ dimensional unit sphere, and $u\in\mathbb{R}^+$. By Theorem 2.1.13 of Muirhead (1982), $d\mu_q(x)=u^{q-1}d\nu_q(x^s)d\mu_1(u)$. Further, for $y\in\mathbb{R}$, consider the change of variable $y=y^su$ with $u\in\mathbb{R}^+$ and $y^s\in\mathbb{R}$, so that $d\mu_1(y)=ud\mu_1(y^s)$. We thus can write the joint density of $(X^s,Y^s,U)$ with respect to $\nu_q\times\mu_1\times\mu_1$ as

$$(2\pi)^{-(q+1)/2}|\Sigma|^{-1/2}\exp[-\tfrac{1}{2}\begin{pmatrix}x^su\\y^su\end{pmatrix}'\Sigma^{-1}\begin{pmatrix}x^su\\y^su\end{pmatrix}]u^q$$

and the marginal density of $W^s=(X^{s\prime},Y^s)'$ with respect to $\nu_q\times\mu_1$ is

$$\begin{aligned}
&(2\pi)^{-(q+1)/2}|\Sigma|^{-1/2}\int_0^\infty u^q\exp[-\tfrac{1}{2}u^2(w^{s\prime}\Sigma^{-1}w^s)]d\mu_1(u)\\
=\;&(2\pi)^{-(q+1)/2}|\Sigma|^{-1/2}\tfrac{1}{2}\int_0^\infty t^{(q-1)/2}\exp[-\tfrac{1}{2}t(w^{s\prime}\Sigma^{-1}w^s)]d\mu_1(t)\\
=\;&(2\pi)^{-(q+1)/2}|\Sigma|^{-1/2}\tfrac{1}{2}\Gamma(\tfrac{q+1}{2})2^{(q+1)/2}\left(w^{s\prime}\Sigma^{-1}w^s\right)^{-(q+1)/2}\\
=\;&\tfrac{1}{2}\pi^{-(q+1)/2}|\Sigma|^{-1/2}\Gamma(\tfrac{q+1}{2})\left(w^{s\prime}\Sigma^{-1}w^s\right)^{-(q+1)/2}
\end{aligned}$$

where the second equality follows from the form of the Gamma density function.

## 9.4 More Details on the Frequentist Prediction Set

### 9.4.1 Approximate Least Favorable Distributions

The problem that underlies the frequentist set is

$$V^* = \inf_A E_\Gamma\left(\text{vol}(A(X^s))\right) \text{ subject to } \inf_{\theta\in\Theta} P_\theta(Y^s \in A(X^s)) \geq 1-\alpha \tag{23}$$

where $\alpha > 0$, $E_\Gamma$ denotes expectation with respect to $f^\Gamma_{X^s}$ and $P_\theta$ denotes that the probability is computed using the joint distribution of $X^s$ and $Y^s$ given $\theta$. If there exists a finite interval $B \subset \mathbb{R}$ such that $\inf_{\theta\in\Theta} P_\theta(Y^s \in B) \geq 1-\alpha$, then $V^*$ must be finite. Let $\Omega$ be a probability distribution with support in $\Theta$, and let $P_\Omega$ be the probability distribution induced by first drawing $\theta$ from $\Omega$, and then drawing $(X^s, Y^s)$ given that $\theta$. Let $A_\Omega$ solve the problem

$$\min_A E_\Gamma\left(\text{vol}(A(X^s))\right) \text{ subject to } P_\Omega(Y^s \in A(X^s)) \geq 1-\alpha. \tag{24}$$

Notice that $\min_A E_\Gamma\left(\text{vol}(A_\Omega(X^s))\right) \leq V^*$ because any $A$ that satisfies the coverage constraint in (23) also satisfies the constraint in (24). Thus, if $\Lambda$ is a distribution for $\theta$, $A_\Lambda$ is the set that solves (24), and if $\inf_{\theta\in\Theta} P_\theta(Y^s \in A_\Lambda(X^s)) \geq 1-\alpha$, then $A_\Lambda$ solves (23). Such a distribution $\Lambda$ is the least favorable distribution for this problem (cf. Theorem 3.8.1 of Lehmann and Romano (2005) and its proof for the corresponding results in hypothesis testing).

The solution to (24) is of the form (11), that is $A_\Omega(x^s) = \{y^s : f^\Omega_{(Y^s,X^s)}(y^s, x^s) > \text{cv}\, f^\Gamma_{X^s}(x^s)\}$, where $\text{cv} > 0$ is chosen so that $P_\Omega(Y^s \in A_\Omega(X^s)) = 1-\alpha$ (such a cv always exists as long as $f^\Omega_{(Y^s,X^s)}(Y^s, X^s)/f^\Gamma_{X^s}(X^s)$ is a continuous random variable under $f^\Omega_{(Y^s,X^s)}$, as is the case in our application). To see this, note that any $A$ is equivalently characterized by the 'test'-function $\varphi : \mathbb{R}^q \times \mathbb{R} \mapsto \{0,1\}$ defined via $\varphi(y^s, x^s) = \mathbf{1}[y^s \in A(x^s)]$. In this notation, $E_\Gamma\left(\text{vol}(A(X^s))\right) = \int\int f^\Gamma_{X^s}(x^s)\varphi(y^s, x^s)d\nu_q(x^s)d\mu_1(y^s) = \int \varphi(w^s)f_1(w^s)d\lambda_{q,1}(w^s)$, and $P_\Omega(Y^s \in A(X^s)) = \int\int f^\Omega_{W^s}(x^s, y^s)\varphi(y^s, x^s)d\nu_q(x^s)d\mu_1(y^s) = \int \varphi(w^s)f_0(w^s)d\lambda_{q,1}(w^s)$, where $d\lambda_{q,1}(w^s) = d\nu_q(x^s) \times d\mu_1(y^s)$, $f_1(w^s) = f^\Gamma_{X^s}(x^s)$ and $f_0(w^s) = f^\Omega_{W^s}(x^s, y^s)$. Thus, the program (24) is equivalent to the problem of finding the best 'test' that rejects with probability at least $1-\alpha$ when the density of $W^s$ is $f_0$, and minimizes the "rejection probability" when the density of $W^s$ is $f_1$. This latter density is not integrable, but the solution still has to be of the Neyman-Pearson form (11), as can be seen by the very argument that proves the Neyman-Pearson Lemma: Let $\varphi^*$ correspond to $A_\Omega$, and let some other $\varphi : \mathbb{R}^{q+1} \mapsto \{0,1\}$ satisfy $\int \varphi f_0 d\lambda_{q,1} \geq 1-\alpha$ (we drop $w^s$ as the dummy variable of integration for convenience). Then

$$\begin{aligned} 0 &\leq \int (\varphi^* - \varphi)(f_0 - \text{cv}\, f_1)d\lambda_{q,1} \\ &\leq \text{cv}\Big(\int \varphi f_1 d\lambda_{q,1} - \int \varphi^* f_1 d\lambda_{q,1}\Big) \end{aligned}$$

where the first inequality follows from the definition of $\varphi^*$ and the second from $1-\alpha = \int \varphi^* f_0 d\lambda_{q,1} \leq \int \varphi f_0 d\lambda_{q,1}$.

WAV minimizing bet-proof sets as suggested by Müller and Norets (2012) described in the main text correspond to the program (24) with the additional constraint that the prediction sets $A$ must contain, for all realizations of $x^s$, the set $[q^{\Gamma}_{\alpha/2}(x^s); q^{\Gamma}_{1-\alpha/2}(x^s)]$. The argument for the constrained optimality of the set (12) follows from the same line of reasoning as above.

Following Elliott, Müller, and Watson (2013) we use numerical methods to approximate the least favorable distribution $\Lambda$ by a distribution $\tilde{\Lambda}$. In the construction of $\tilde{\Lambda}$, we solve (24) a little more stringently (that is, the constraint is $P_{\tilde{\Lambda}}(Y^s \in A(X^s)) \geq 1-\alpha+\epsilon$ for some small $\epsilon > 0$) to facilitate $\inf_{\theta\in\Theta} P_\theta(Y^s \in A_{\tilde{\Lambda}}(X^s)) \geq 1-\alpha$. Since $\min_A E_\Gamma\left(\text{vol}(A(X^s))\right) \leq V^*$ for any $\Omega$, we can numerically assess the cost of $\epsilon > 0$ by comparing $E_\Gamma\left(\text{vol}(A_{\tilde{\Lambda}}(X^s))\right)$ to the lower bound on $V^*$ induced by the approximate least favorable $\tilde{\Lambda}$ with $\epsilon = 0$.

### 9.4.2 Numerical Approximations

*Determination of approximate least favorable distribution:* The algorithm consists of two main steps: (I) determination of a candidate least favorable distribution $\tilde{\Lambda}$; (II) numerical check of $\inf_{\theta\in\Theta} P_\theta(Y^s \in A_{\tilde{\Lambda}}(X^s)) \geq 1-\alpha$.

Both steps are based on a discrete grid of values for $\theta$, with the grid $\Theta_{II}$ for step (II) much finer than the grid $\Theta_I$ for step (I). In particular, with $\Delta_I = \{-0.4, -0.2, \ldots, 1.2\}$, $\Delta_{II} = \{-0.4, -0.3, \ldots, 1.4\}$, $\tilde{B}_I = \{0, 0.01, 0.05, 0.2, 0.5, 1, 2, 5, 20, 80\}$ and $\tilde{B}_{II} = \{0, 0.004, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.5, 1, 1.5, 2, 3, 5, 10, 20, 50, 80, 200\}$, $C_I = \{0, 0.2, 0.5, 2, 10, 80\}$, $C_{II} = \{0, 0.05, 0.2, 0.5, 2, 5, 10, 40, 80, 200\}$, $\Theta_I = \{(b, 0, d)' : d \in \Delta_I, (8\pi)^{2d} b^2 \in \tilde{B}_I\} \cup \{(0, c, d)' : d \in \Delta_I, c \in C_I\} \cup \{(b, c, 1.4)' : c \in C_I, ((8\pi)^2 + c^2)^d b^2 \in \tilde{B}_I\}$ and $\Theta_{II} = \{(b, c, d)' : c \in C_{II}, d \in \Delta_{II}, ((8\pi)^2 + a^2)^d b^2 \in \tilde{B}_{II}\}$. For each $\theta$, $P_\theta(Y^s \in A_{\tilde{\Lambda}}(X^s))$ is approximated via Monte Carlo integration using an importance sampling scheme: For $\Theta_g \in \{\Theta_I, \Theta_{II}\}$ with $|\Theta_g|$ elements, note that

$$P_{\theta_0}(Y^s \in A_{\tilde{\Lambda}}(X^s)) = |\Theta_g|^{-1} \sum_{\theta\in\Theta_g} E_\theta(\text{LR}_{\theta_0}(W^s)\mathbf{1}[Y^s \in A_{\tilde{\Lambda}}(X^s)]) \tag{25}$$

where $E_\theta$ denoted expectation with respect to $f^\theta_{W^s}$, $\text{LR}_{\theta_0}(W^s) = f^{\theta_0}_{W^s}(W^s)/(|\Theta_g|^{-1}\sum_{\theta\in\Theta_g} f^\theta_{W^s}(W^s))$, and $f^\theta_{W^s}$ is the joint distribution of $W^s = (X^{s\prime}, Y^s)'$ as in (9) with $\Sigma = \Sigma(\theta)$ as determined by $\theta$. The Monte Carlo approximation replaces each $E_\theta$ in (25) by an average over 2000 and 400 independent draws from $f^\theta_{W^s}$ in steps (I) and (II), respectively. The fewer draws for step (II) are compensated by the larger number of points in $\Theta_{II}$. The resulting Monte Carlo standard errors are less than 1.1% for all $P_{\theta_0}(Y^s \in A_{\tilde{\Lambda}}(X^s))$ if $\alpha = 0.5$, and 0.7% for $\alpha = (0.1, 0.2)$. Draws from $f^\theta_{W^s}$ are generated via $W^s = W/\sqrt{X'_{1:q}X_{1:q}}$ with $W = (X'_{1:q}, Y)' \sim \mathcal{N}(0, \Sigma(\theta))$.

We now discuss Step I of the algorithm in the more complicated case where $A(X^s)$ is restricted to contain the Bayes set $[q^\Gamma_{\alpha/2}(X^s); q^\Gamma_{1-\alpha/2}(X^s)]$, that is $A = A^{MN}$ in the notation of Section 4. The prior $\Gamma$ is approximated by the discrete prior that puts equal mass on all points of the form $(d, 0, 0)'$, $d \in \{-0.4, -0.2, \ldots, 1.4\}$. Let $\gamma(\theta)$ be the corresponding weights on $\Theta_I$. We initially draw 2000 independent $W^s$ under each $\theta_i \in \Theta_I$, for a total of $2000|\Theta_I|$ draws. For each draw $w^s = (x^{s\prime}, y^s)'$, we first determine the value of the indicator $I(w^s) = \mathbf{1}[y_s \notin [q^\Gamma_{\alpha/2}(x^s); q^\Gamma_{1-\alpha/2}(x^s)]]$. The event $I(w^s)$ is equivalent to the cdf of the posterior distribution of $Y^s$ given $X^s = x^s$, evaluated at $y^s$, $F^\Gamma_{Y^s|X^s}(y^s|x^s)$, to take on a value outside $[\alpha/2, 1-\alpha/2]$, and $F^\Gamma_{Y^s|X^s}(y_s|x_s)$ is given by

$$F^\Gamma_{Y^s|X^s}(y^s|x^s) = \int_{-\infty}^{y_s} \sum_{\theta\in\Theta_I} \pi(\theta) \frac{f^\theta_{W^s}((x^{s\prime}, \tilde{y}^s)')}{f^\theta_{X^s}(x^s)} d\tilde{y}^s \tag{26}$$

where $\pi(\theta_0) = \gamma(\theta_0) f^{\theta_0}_{X^s}(x^s) / (\sum_{\theta\in\Theta_I} \gamma(\theta) f^\theta_{X^s}(x^s))$ is the posterior probability of the data being drawn from parameter value $\theta_0$. The integral in (26) is evaluated using Gaussian quadrature with 100 points, after the change of variables $\tilde{y}^s = \tilde{\sigma}\Phi^{-1}(u)$, where $\Phi^{-1}$ is quantile function of a standard normal and $\tilde{\sigma} = 4\sigma_Y / \sqrt{\sum_{i=1}^q \sigma^2_{X,i}}$, with $\sigma^2_{X,i}$ and $\sigma^2_Y$ the diagonal elements of the $\Sigma(\theta)$ from which $w^s$ was drawn.

Now for a pair of approximately least favorable distribution $\tilde{\Lambda}$ and critical value cv, let $\tilde{\lambda}(\theta)$ be the corresponding nonnegative weight function on $\Theta_I$ with total mass $1/\,\text{cv}$. The event $Y^s \in A_{\tilde{\Lambda}}(X^s)$ in (12) is then equivalent to $\sum_{\theta\in\Theta_I} \tilde{\lambda}(\theta) f^\theta_{W^s}(W^s) \geq \sum_{\theta\in\Theta_I} \pi(\theta) f^\theta_{X^s}(X^s) I(W^s)$. Thus $Q(\theta_0, \tilde{\lambda}) = P_{\theta_0}(Y^s \in A_{\tilde{\Lambda}}(X^s)) = P_{\theta_0}(\sum_{\theta\in\Theta_I} \tilde{\lambda}(\theta) f^\theta_{W^s}(W^s) \geq \sum_{\theta\in\Theta_I} \pi(\theta) f^\theta_{X^s}(X^s) I(W^s))$. Note that the least favorable distribution $\Lambda$ cannot put any mass on points $\theta$ where $P_\theta(Y^s \in A_{\tilde{\Lambda}}(X^s)) > 1-\alpha$, as this would imply that also $P_\Lambda(Y^s \in A(X^s)) > 1-\alpha$, which contradicts the solution to (24) discussed above. Thus, we seek values for $\tilde{\lambda}$ that satisfy

$$\begin{aligned} Q(\theta, \tilde{\lambda}) &= 1-\alpha+\epsilon \quad \text{if } \tilde{\lambda}(\theta) > 0 \\ Q(\theta, \tilde{\lambda}) &\geq 1-\alpha+\epsilon \quad \text{if } \tilde{\lambda}(\theta) = 0. \end{aligned}$$

An appropriate solution is determined iteratively: set $\tilde{\lambda}_{(i)}(\theta) = \exp(\eta_{(i)}(\theta))$, where $\eta_{(i)}(\theta) \in \mathbb{R}$ for all $\theta \in \Theta_I$. Start with $\eta_{(0)}(\theta) = -9$ for all $\theta \in \Theta_I$. Then set

$$\eta_{(i)}(\theta) = \eta_{(i-1)}(\theta) - \kappa(Q(\theta, \tilde{\lambda}_{(i-1)}) - (1-\alpha+\epsilon)) \text{ for all } \theta \in \Theta_I$$

that is, decrease (increase) the value of $\tilde{\lambda}_{(i)}$ proportional to the degree of overcoverage (undercoverage) at $\theta \in \Theta_I$. This process is run with $\kappa = 2$ for 4000 iterations.

The value of the slack parameter $\epsilon$ is $(0.003, 0.005, 0.01)$ for $\alpha = (0.1, 0.2, 0.5)$ for $q \in \{6, 9, 12, 24\}$, and equal to $(0.005, 0.02, 0.02)$, for $q = 48$. For $q = 48$, more slack is necessary to ensure coverage under the finer grid $\Theta_{II}$. This additional slack leads to prediction sets

that are no more than 1.0-2.5% and 1.0-5.0% longer than any set that achieves uniform coverage for $q \leq 24$ and $q = 48$, respectively, depending on the horizon $r$. Step (I) of the algorithm is run for $r \in R_I = \{0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4\}$, and step (II) for $r \in \{0.05, 0.1, 0.3, 0.6, 1, 1.4\}$. This requires a considerable number of calculations (which take approximately 72 hours on currently available desktop computers, with most of the time spent on step (II)). Numerical calculations suggest that all sets $A_{\tilde{\Lambda}}^{MN}$ are intervals. In the empirical work, prediction intervals for values of $r$ not included in $R_I$ are computed via linear interpolation.

*Computation of $\Sigma$ in bcd-model:*
As discussed in the main text, the elements of $\Sigma$ are given by $2\int_0^\infty S(\omega)w_{jk}(\omega)d\omega$, where $w_{jk}(\omega) = \text{Re}[\left(\int_0^{1+r} g_j(s)e^{-i\omega s}ds\right)\left(\int_0^{1+r} g_k(s)e^{i\omega s}ds\right)]$ and $S(\omega)$ is available in closed form for a given $\theta$. The functions $w_{jk}$ are determined in closed form using computer algebra. The integral $\int_0^\infty S(\omega)w_{jk}(\omega)d\omega$ is then split into the sum of the corresponding integral over $(0, e^{-1})$, and the interval $[e^{-1}, \infty)$, which are evaluated using Gaussian quadrature with 1,000 points each. The first integral uses the transformation of variables $\omega = \exp(-(10u+1)^2)$, $u \in [0,1]$, and relies on a fourth order Taylor approximation of $w_{jk}$ around $\omega = 0$, and the second integral is approximated by Gaussian quadrature on the interval $[e^{-1}, 10\pi q]$.

# Data Appendix

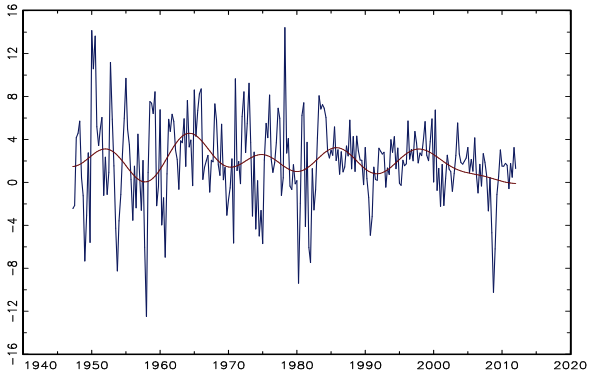
| Series | Sample period | Description | Sources and Notes |
|---|---|---|---|
| ***Post- WWII Quarterly Series*** | | | |
| GDP | 1947:Q1-2012:Q1 | Real GDP (Billions of Chained 2005 Dollars) | FRED Series: GDPC96 |
| Consumption | 1947:Q1-2012:Q1 | Real Personal Consumption Expenditures (Billions of Chained 2005 Dollars) | FRED Series PCECC96 |
| Total Factor Productivity | 1947:Q2-2012:Q1 | Growth Rate of TFP | From John Fernald's Web Page: Filename Quarterly)TFP.XLSX, series name DTFP. The series is described in Fernald (2012) |
| Labor Productivity | 1947:Q1-2012:Q1 | Output per hour in the non-farm business sector | FRED Series: OPHNFB |
| Population | 1947:Q1-2012:Q1 | Population (mid Quarter) (Thousands) | Bureau of Economic Analysis NIPA Table 7.1. Adjusted for an outlier in 1960:Q1 |
| Prices: PCE Deflator | 1947:Q1-2012:Q1 | PCE deflator | FRED Series: PCECTPI |
| Inflation (CPI) | 1947:Q1-2012:Q1 | CPI | FRED Series: CPIAUCSL. The quarterly price index was computed as the average of the monthly values |
| Inflation (CPI, Japan) | 1960:Q1-2012:Q1 | CPI for Japan | FRED Series: CPI_JPN. The quarterly price index was computed as the average of the monthly values |
| Stock Returns | 1947:Q1-2012:Q1 | CRSP Real Value-Weighted 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 shown are 400×the logarithm of gross quarterly real returns. |
| ***Longer Span** Data **Series*** | | | |
| Real GDP | 1900-2011 | Real GDP (Billions of Chained 2005 Dollars) | 1900-1929: Carter, Gartnter, Haines, Olmstead, Sutch, and Wright (2006), Table Ca9: Real GDP in $1996<br>1929-2011: BEA Real GDP in $2005. Data were linked in 1929 |
| Real Consumption | 1900-2011 | Real Personal Consumption Expenditures (Billions of Chained 2005 Dollars) | 1900-1929: Carter, Gartnter, Haines, Olmstead, Sutch, and Wright (2006), Table Cd78 Consumption expenditures, by type, Total, $1987<br>1929-2011: BEA Real Consumption in $2005. Data were linked in 1929 |
| Population | 1900-2011 | Population | 1900-1949: Carter, Gartnter, Haines, Olmstead, Sutch, and Wright (2006), Table Aa6 (Population Total including Armed Forces Oversees), 1917-1919 and 1930-1959; Table Aa7 (Total, Resident) 1900-1916 and 1920-1929)<br>1950-2011: Bureau of the Census (Total, Total including Armed Forces Oversees), |
| Inflation (CPI) | 1913-2011 | Consumer Price Index | BLS Series: CUUR0000SA0 |
| Stock Returns | 1926:Q1-2012:Q1 | CRSP Real Value-Weighted Returns | Described above |

Generally, the data used in the paper are growth rates computed as differences in logarithms in percentage points at annual rate ($400\times\ln(X_t/X_{t-1})$ if $X$ is measured quarterly and $100\times\ln(X_t/X_{-1})$ if $X$ is measured annually). The exceptions are stock returns which are $400\times X_t$, where $X_t$ is the gross quarterly real return (that is, $X_t = 1+R_t$, where $R_t$ is the net return), and TFP which is reported in percentage points at annual rate in the source data.

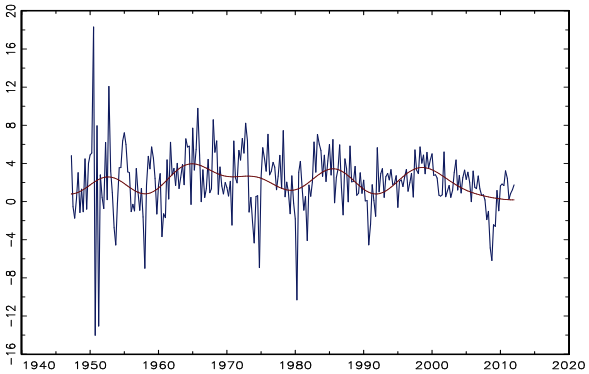
## Additional Figures

Figure A.1: Time Series and Low-Frequency Components

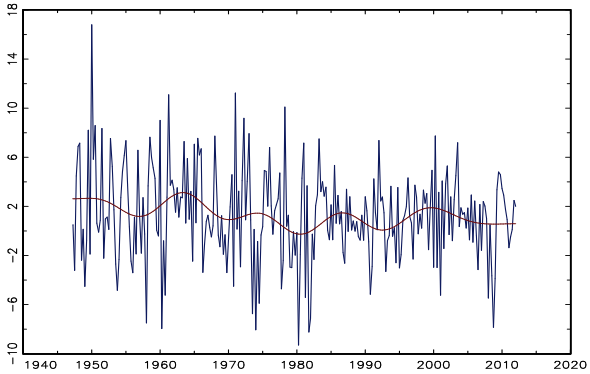
A. Real Per-Capita GDP



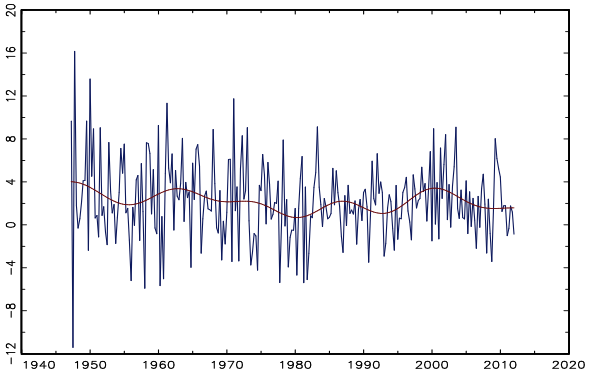

B. Real Consumption Per Capita



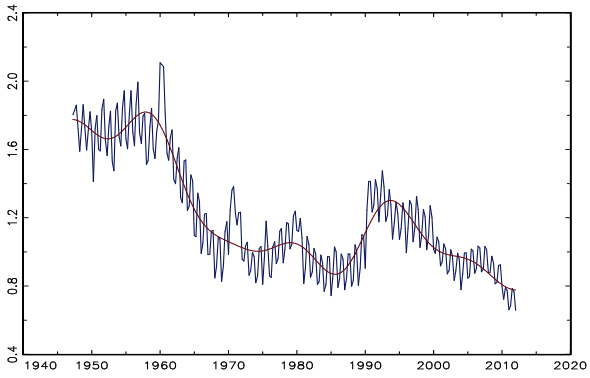

C. Total Factor Productivity




D. Labor Productivity




E. Population




F. Inflation (PCE)

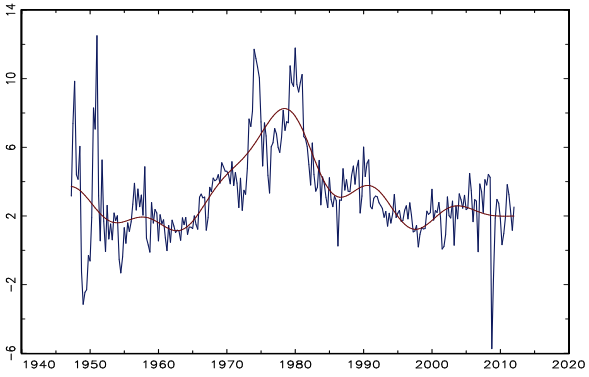
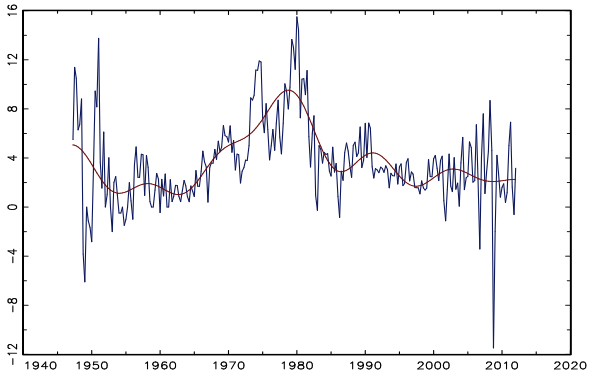
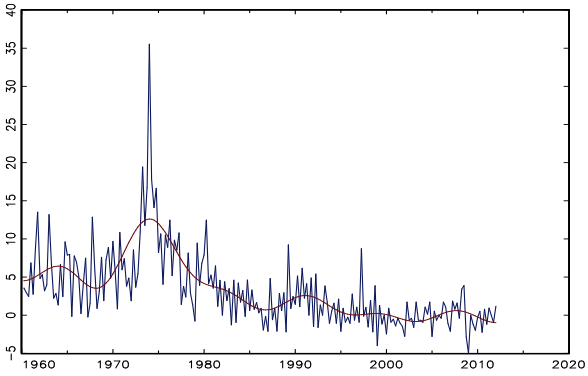


Figure A.1: Time Series and Low-Frequency Components
(Continued)

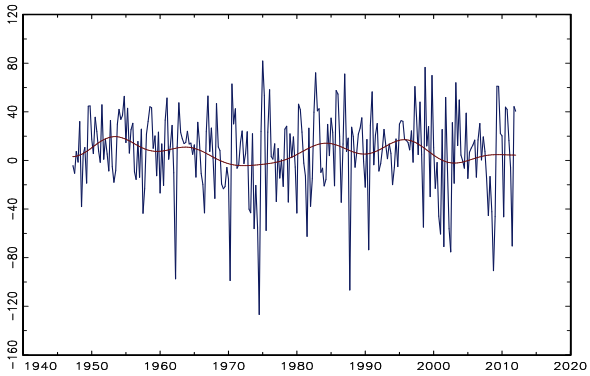
G. Inflation (CPI)



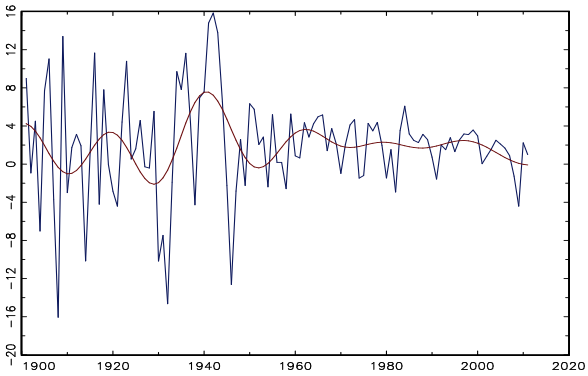

H. Inflation (CPI, Japan)



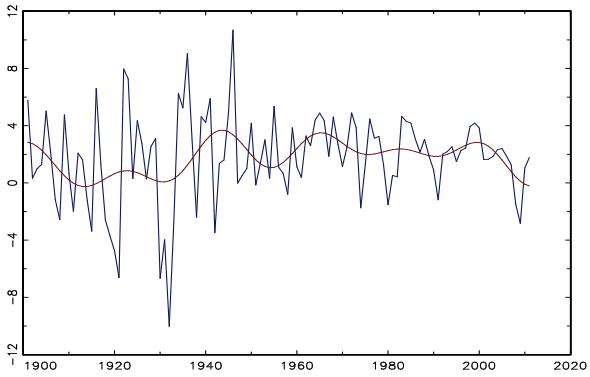

I. Stock Returns




J. Real GDP per capita (Annual)




K. Real Cons. per capita (annual)




L. Population (Annual)

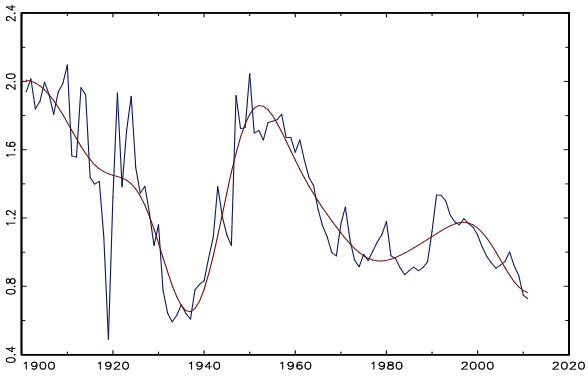
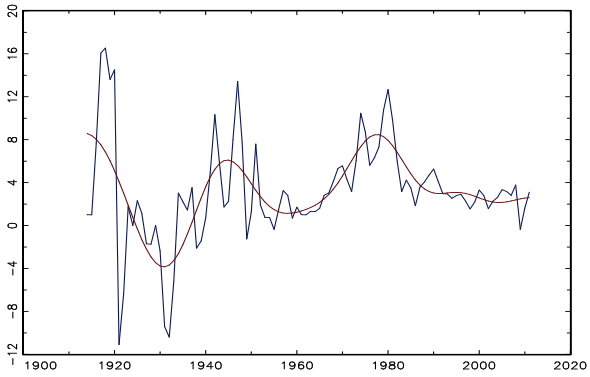


Figure A.1: Time Series and Low-Frequency Components
(Continued)

M. Inflation (CPI, Annual)




N. Stock Returns

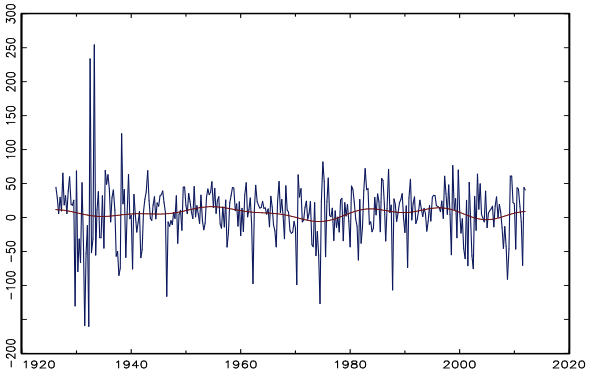
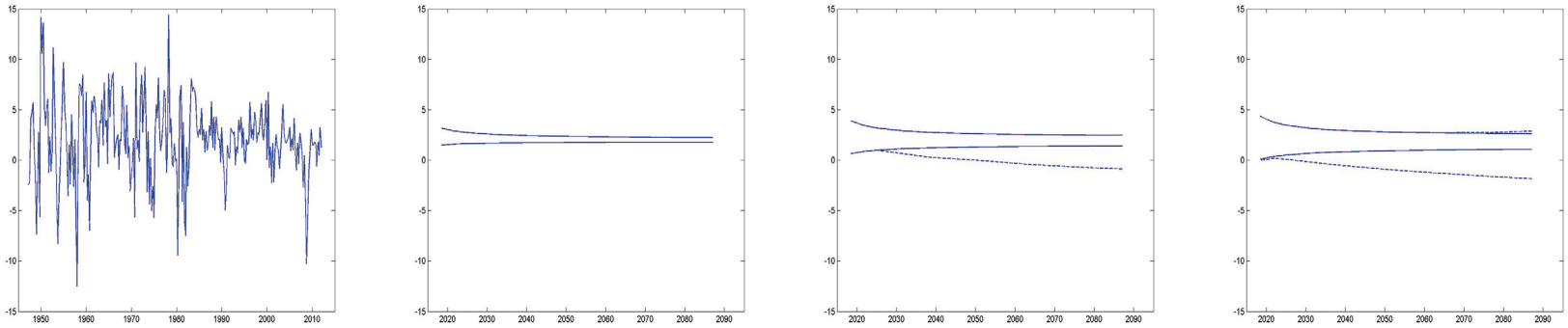


Figure A.2: Prediction Sets
(Bayes: Solid and Frequentist: Dashed)

A. Real Per-Capita GDP

(i) Data (ii) 50% (iii) 80% (iv) 90%




B. Real Consumption Per Capita

(i) Data (ii) 50% (iii) 80% (iv) 90%

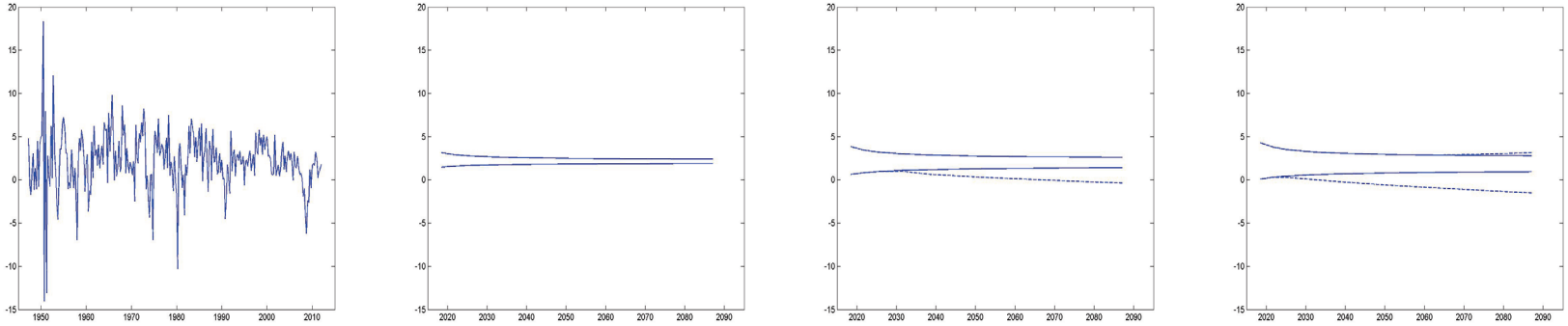
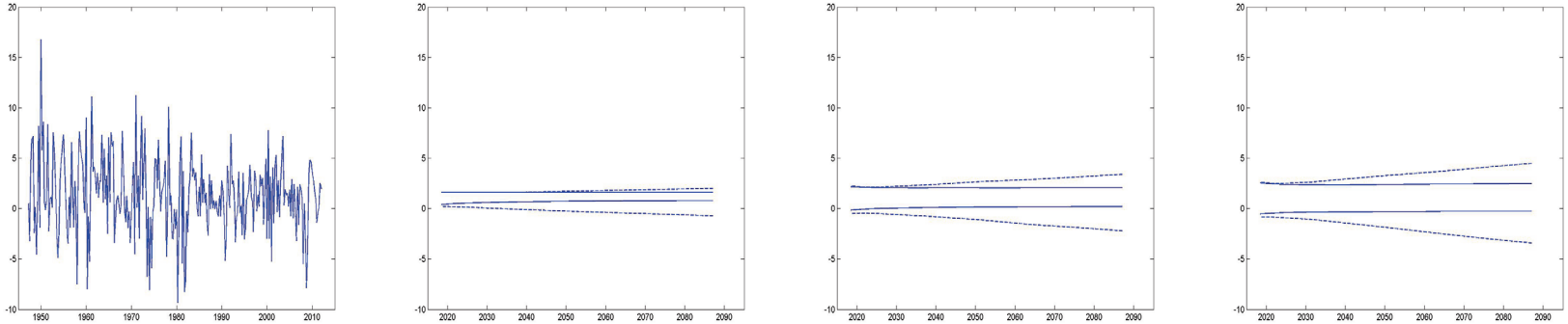


Figure A.2: Prediction Sets (Continued)
(Bayes: Solid and Frequentist: Dashed)

C. Total Factor Productivity

(i) Data (ii) 50% (iii) 80% (iv) 90%



D. Labor Productivity

(i) Data (ii) 50% (iii) 80% (iv) 90%

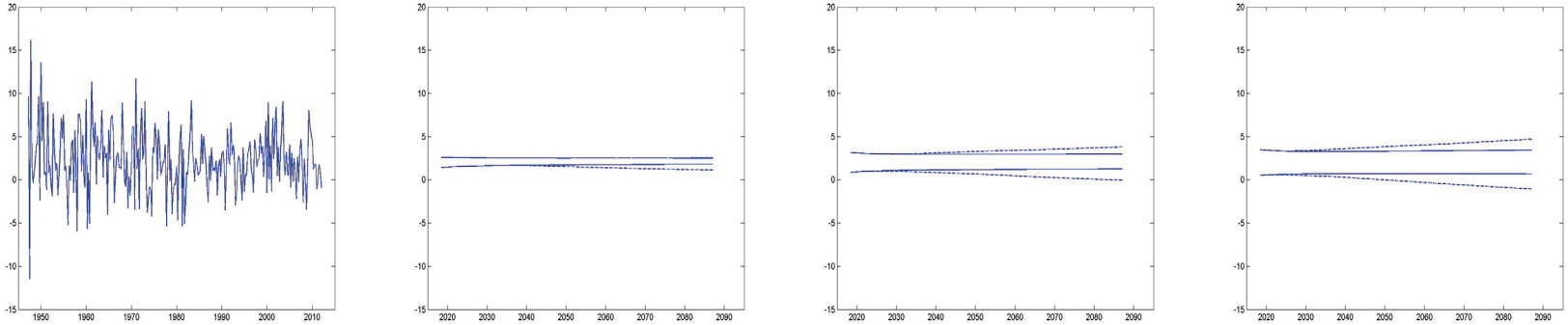
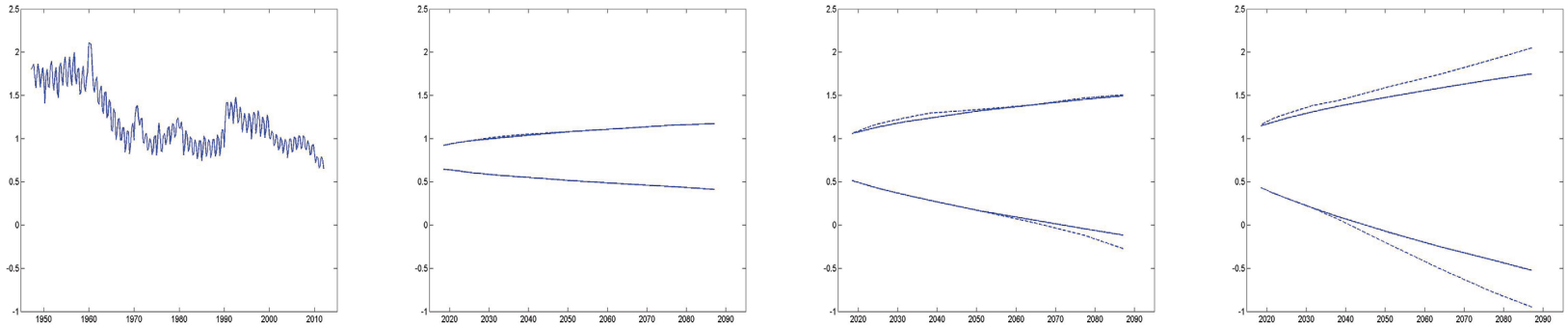


Figure A.2: Prediction Sets (Continued)
(Bayes: Solid and Frequentist: Dashed)

E. Population

(i) Data (ii) 50% (iii) 80% (iv) 90%



F. Inflation (PCE)

(i) Data (ii) 50% (iii) 80% (iv) 90%

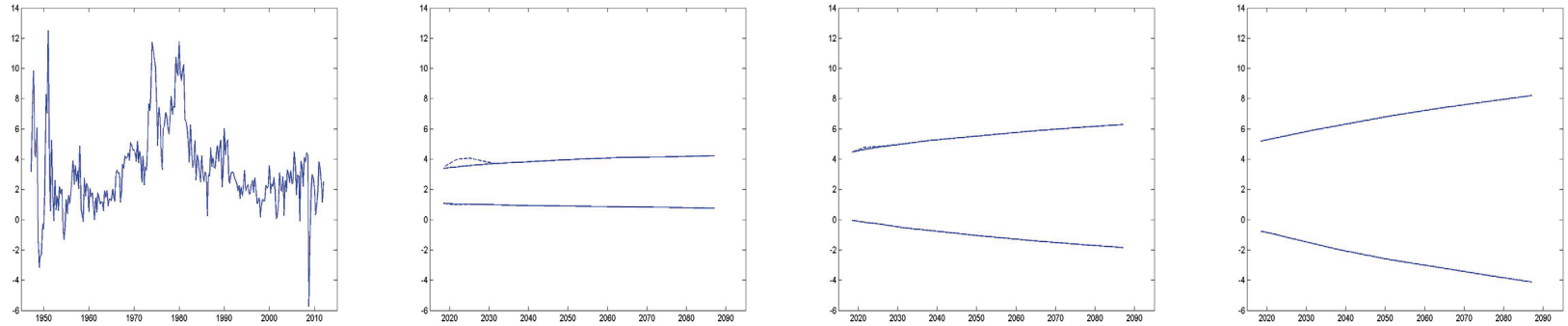
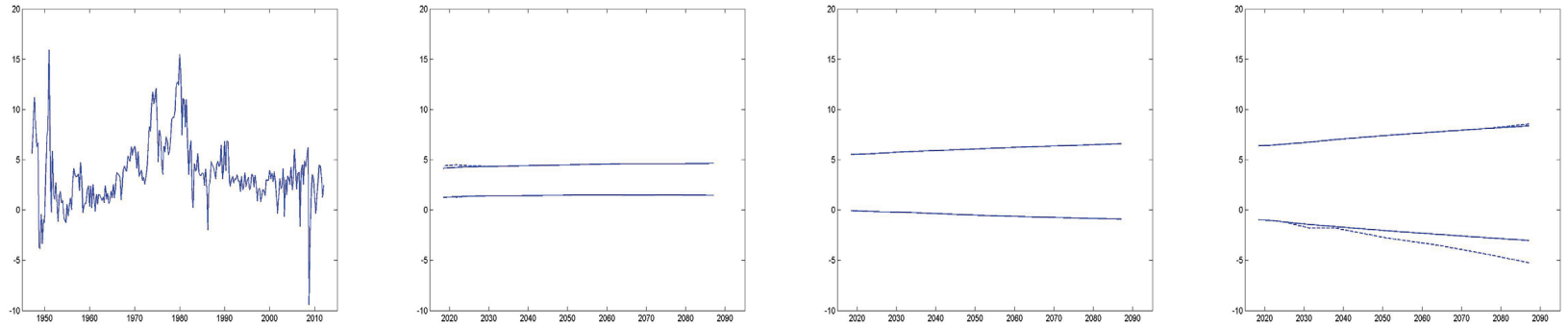


Figure A.2: Prediction Sets (Continued)
(Bayes: Solid and Frequentist: Dashed)

G. Inflation (CPI)

(i) Data (ii) 50% (iii) 80% (iv) 90%



H. Inflation (CPI, Japan)

(i) Data (ii) 50% (iii) 80% (iv) 90%

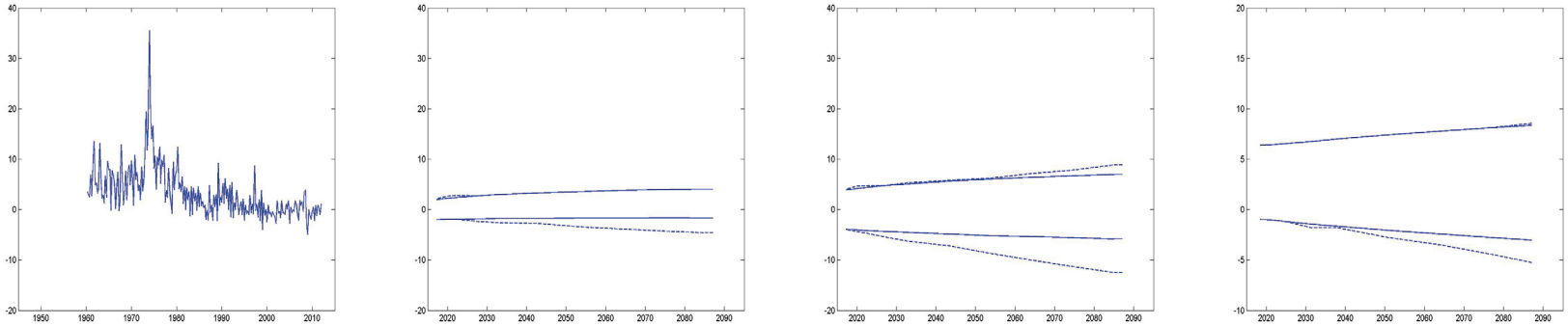
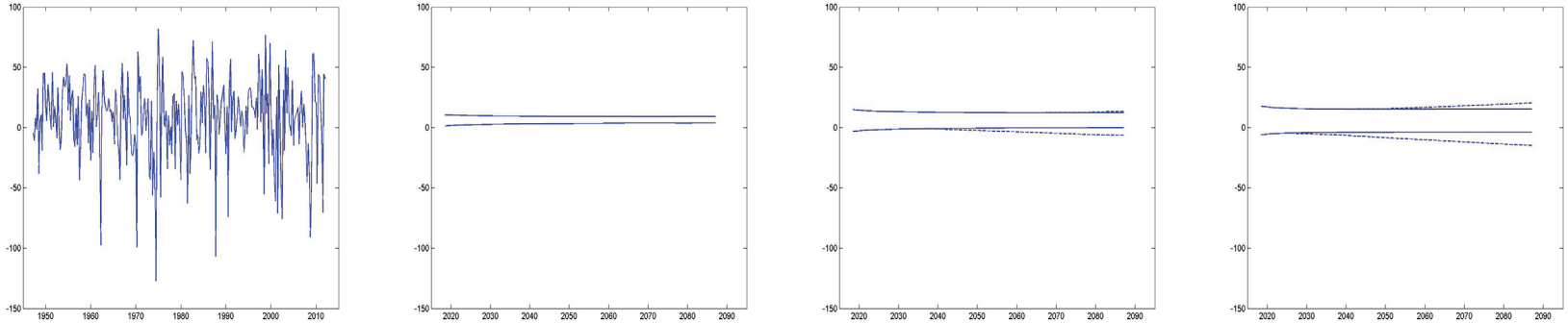


Figure A.2: Prediction Sets (Continued)
(Bayes: Solid and Frequentist: Dashed)

I. Stock Returns

(i) Data (ii) 50% (iii) 80% (iv) 90%




J. Real GDP per capita (Annual)

(i) Data (ii) 50% (iii) 80% (iv) 90%

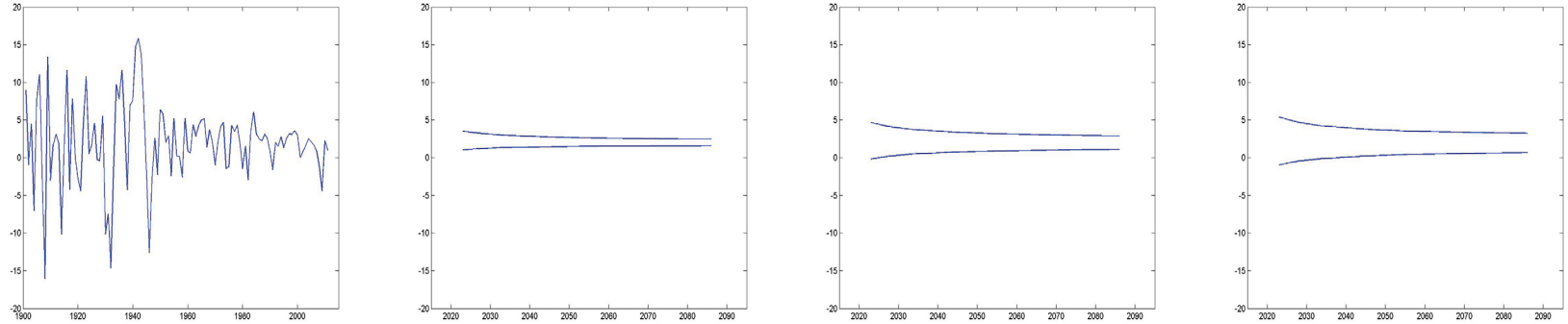
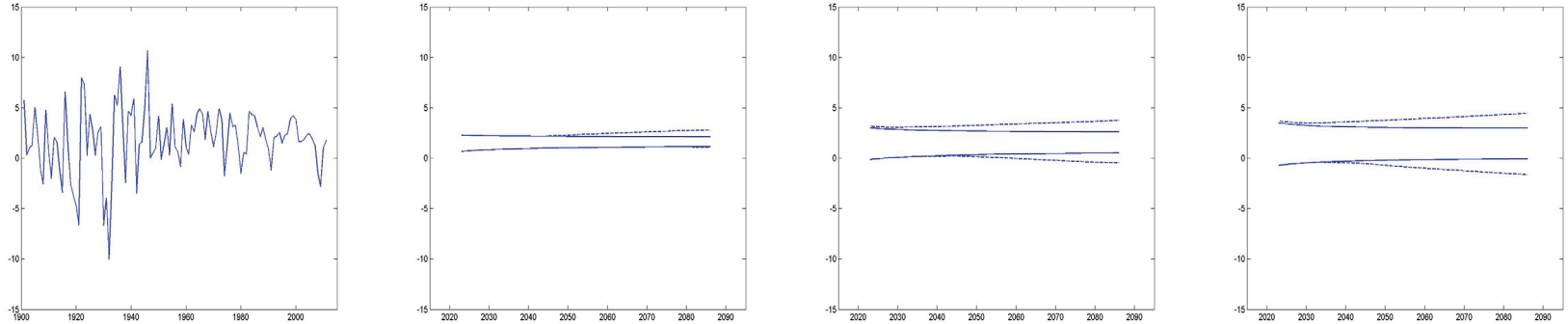


Figure A.2: Prediction Sets (Continued)
(Bayes: Solid and Frequentist: Dashed)

K. Real Cons. per capita (annual)

(i) Data (ii) 50% (iii) 80% (iv) 90%




L. Population (Annual)

(i) Data (ii) 50% (iii) 80% (iv) 90%

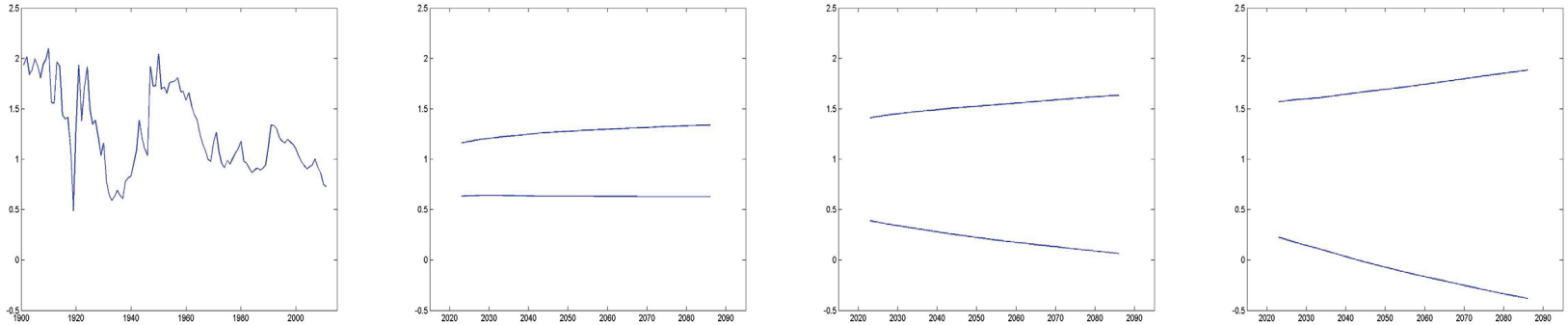
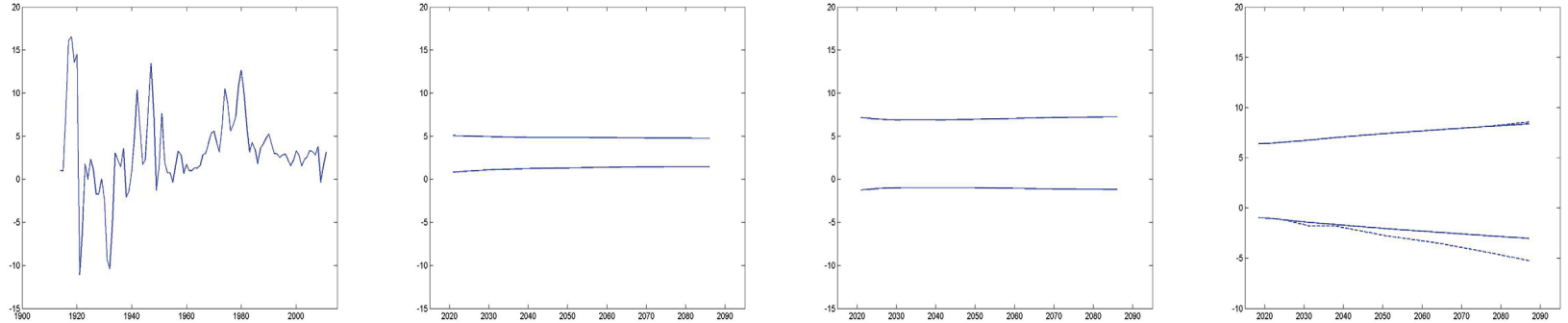


Figure A.2: Prediction Sets (Continued)
(Bayes: Solid and Frequentist: Dashed)

M. Inflation (CPI, Annual)

(i) Data (ii) 50% (iii) 80% (iv) 90%



N. Stock Returns

(i) Data (ii) 50% (iii) 80% (iv) 90%