Measuring Uncertainty about Long-Run Predictions*

Ulrich K. Müller and Mark W. Watson
Princeton University
Department of Economics
Princeton, NJ, 08544
February 2013 (Revised September 2015)

Abstract

Long-run forecasts of economic variables play an important role in policy, planning, and portfolio decisions. We consider forecasts of the long-horizon average of a scalar variable, typically the growth rate of an economic variable. The main contribution is the construction of prediction 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 prediction sets for 10 to 75 year average growth rates of U.S. real per-capita GDP and consumption, productivity, price level, stock prices and population.

JEL classification: C22, C53, E17

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

*We thank Frank Diebold, Graham Elliott, Bruce Hansen, James Stock, Jonathan Wright, three referees, and the Editor, Stéphane Bonhomme, for useful comments and advice. Support was provided by the National Science Foundation through grants SES-0751056 and SES-1226464. Replication files and a supplementary appendix is available at http://www.princeton.edu/~mwatson.

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-horizon returns.[3]

Let x_t denote a time series, such as the inflation rate, the growth rate of real GDP or the return on a portfolio of stocks. Sample data on x_t are available for $t = 1, ..., T$, say 1947-2014. 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 $T + h$, say the 32-year horizon 2014-2046. 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%). This is a long-horizon problem, since the horizon h is large relative to the number of available observations T (in the example, $r = h/T \approx 0.5$).

We structure the problem so that the coverage probability can be calculated using asymptotic approximations based on a 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 covariance matrix of the normal distribution in the large-sample

[1] See Congressional Budget Office (2005).

[2] See Fleckenstein, Longstaff, and Lustig (2013), Hilsher, Raviv, and Reis (2014) and Kitsul and Wright (forthcoming), who use market prices on various inflation-related derivatives to estimate market-based predictive distributions of inflation.

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

problem. We assume that the first differences $\Delta x_t = x_t - x_{t-1}$ are covariance stationary. (Recall that x_t is a series like the growth rate of real GDP, inflation, or asset returns, so this does not rule out stochastic trends in these *growth rates*.) Since we are interested in a long-run prediction ($\overline{x}_{T+1:T+h}$, for h large relative to T) the crucial characteristic of x_t is its (pseudo-) spectrum near frequency zero. The relative paucity of sample information about these low-frequency properties precludes a nonparametric approach. We therefore proceed by constructing a flexible parametric model for the shape of the spectrum near frequency zero that nests the fractional, local-to-unity and local-level forms of long run persistence. The uncertainty about the parameter θ of this model in turn becomes an important component of the uncertainty about $\overline{x}_{T+1:T+h}$.

We use both Bayes and frequentist methods to incorporate this uncertainty in our prediction sets. The Bayes procedure is straightforward: given a prior for the parameter θ, 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. While Bayes sets have many desirable properties, they have the potentially undesirable property of controlling coverage (that is, the probability that the set includes the future value of $\overline{x}_{T+1:T+h}$) only *on average* for values of θ drawn from the prior. Thus in general, coverage will fall short of the nominal level for some values of θ, and the specifics of this undercoverage will depend on the prior used. To address this limitation we robustify the Bayes prediction sets by enlarging them so that they have frequentist properties: the resulting sets provide (possibly conservative) coverage for all values of θ. Using ideas borrowed from Müller and Norets (2012), we do this in a way that minimizes the sets' average expected length.

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 computed from probability models over long horizons, and the literature on this topic is relatively sparse. 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. Two strands of literature study long-run forecast uncertainty for time series that we analyze by constructing series-specific Bayesian models. Pastor and Stambaugh (2012) compute predictive variances of long-run forecasts of stock returns that account for parameter uncertainty. Lee (2011) and Raftery, Li, Sevcíkova, Gerland, and Heilig (2012) study long-run forecasts of population and fertility rates.

The outline of this paper is as follows. Section 2 formalizes the long-horizon prediction problem and discusses the low-frequency summaries of the sample data used in the analysis. This section also introduces two running examples: forecasting the average growth rate of real per-capita GDP and the average level of consumer price index (CPI) inflation in the U.S. over the next 25 years. Section 3 discusses and develops the requisite statistical tools for constructing the long-horizon prediction sets. Two sets of tools are needed. The first is a central limit theorem and associated covariance matrix that yields a large-sample Gaussian version of the prediction problem. The second are methods for constructing Bayes and frequentist prediction sets for this limiting problem. The Bayes procedures are standard; the frequentist procedures are not, and are developed in Section 3.3. Section 4 takes up the important practical problems of parameterizing the covariance matrix in the limiting problem (which involves parameterizing the spectrum of x_t near frequency 0), choosing a prior for the Bayes prediction sets and a related weighting function for the frequentist sets (to obtain a scalar criterion for comparing the efficiency of sets), and choosing the number of low-frequency averages of the sample data to use (which involves a classic trade-off between efficiency and robustness). Taken together, Sections 2-4 develop methods for constructing prediction sets with well-defined large-sample optimality properties; these methods are illustrated using the GDP and inflation running examples throughout these sections. Section 5 uses simulations and pseudo-out-of-sample experiments to evaluate the performance of these sets in small samples. One focus of this analysis is the effect of level and/or volatility "breaks" on the prediction sets. Following this extensive background, Section 6 applies these methods to construct prediction sets spanning up to 75 years for eight U.S. economic time series: the running examples of real GDP and CPI inflation, the rates of growth of per-capita consumption expenditures, total factor and labor productivity, population, stock prices, and an alternative measure of price inflation. Section 7 concludes.

2 The Prediction Problem

Let x_t be the economic variable of interest which is observed for $t = 1, \ldots, T$. The objective is to construct a prediction set, denoted by A, of the average value of x_t from periods $T+1$ to $T+h$,

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

with the property that $P(\overline{x}_{T+1:T+h} \in A) = 1 - \alpha$, where α is a pre-specified constant. The prediction set A is constructed using the sample data for x_t, so that $A = A\left(\{x_t\}_{t=1}^{T}\right)$.[4] We restrict A in two ways. First, we allow A to depend on the sample data only through a small number low-frequency weighted averages of the sample data, and second, we restrict A to be scale and location equivariant. We discuss each of these restrictions in turn.

Cosine transformations of the sample data. Because h is large, the prediction sets involve long-run uncertainty about x_t. It is therefore useful to transform the sample data into weighted averages that capture variability at different frequencies – we will be interested in the weighted averages corresponding to low frequencies. Thus, consider the weighted averages $(\overline{x}_{1:T}, X_T)$, with $\overline{x}_{1:T} = T^{-1}\sum_{t=1}^{T} x_t$, $X_T = (X_T(1), \ldots, X_T(T-1))'$, and where $X_T(j)$ is the jth cosine transformation

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

with $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$ and $\iota_{jT} = (2T/j\pi)\sin(j\pi/2T) \to 1$. The cosine transforms have two properties we will exploit. First, they isolate 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 $j\pi/T$. Second, because the Ψ_j weights add to zero, $X_T(j)$ is invariant to location shifts of the sample, a property we use when we construct equivariant prediction sets.

The $T \times 1$ vector $(\overline{x}_{1:T}, X_T)$ is a nonsingular transformation of the sample data $\{x_t\}_{t=1}^{T}$, but we will construct prediction sets based on a truncated information set that includes only $\overline{x}_{1:T}$ and the first q cosine transforms, $X_{T,1:q} = (X_T(1), X_T(2), ..., X_T(q))'$ and where q is much smaller than $T-1$. Thus, the prediction sets we consider are of the form $A = A(\overline{x}_{1:T}, X_{T,1:q})$, and so rely solely on variability in the data associated with frequencies lower than $q\pi/T$.

[4] Of course, when x_t is the first difference of another variable y_t, so that $x_t = y_t - y_{t-1}$, then forecasts of y_{T+h} can be constructed from forecasts of $\overline{x}_{T+1:T+h}$ using the identity $y_{T+h} = y_T + h\overline{x}_{T+1:T+h}$. Moreover, prediction sets for $\overline{x}_{T+1:T+h}$ and y_{T+h} are readily converted into prediction sets for monotonic transformation of these variables. For example, a prediction set for the average growth rate of real GDP ($\overline{x}_{T+1:T+h}$) yields a prediction set for the log-level of real GDP (y_{T+h}) or the level of real GDP ($\exp(y_{T+h})$).

We compress the sample information into the $q+1$ variables $(\overline{x}_{1:T}, X_{T,1:q})$ 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 theorem 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, 2013) discuss this robustness issue in detail. In Section 4 below we present numerical calculations that quantify the efficiency-robustness trade-off embodied by the choice of q in the long-run prediction problem.

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, as mentioned above, $X_{T,1:q}$ 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

$$Y_T = \overline{x}_{T+1:T+h} - \overline{x}_{1:T} \tag{3}$$

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

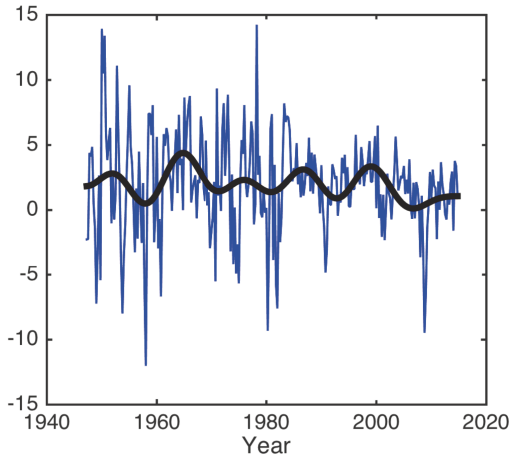
Running examples: Two of the economic time series studied in Section 6 are the growth rate of U.S. real per-capita GDP and the rate of inflation in the U.S. based on the consumer price index. We use these series as running examples to illustrate concepts as they are introduced. Panels (i) in Figure 1 plot the quarterly values of these time series from 1947-2014, along with the low-frequency components of the time series formed as the projection of

[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 $Y_T^s \in A(0, X_{T,1:q}^s)$.

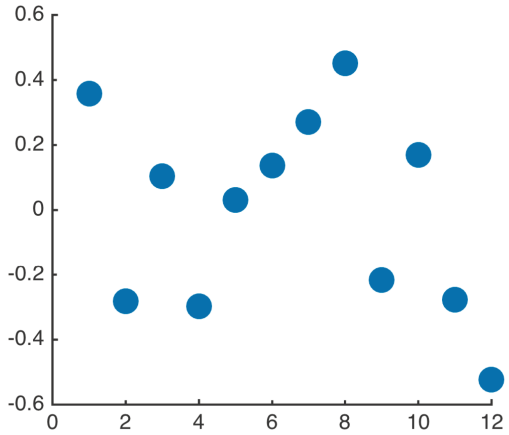
Figure 1: Low-frequency components and cosine transforms

(a) Growth rate of real per-capita GDP

(i) Series and low-frequency component

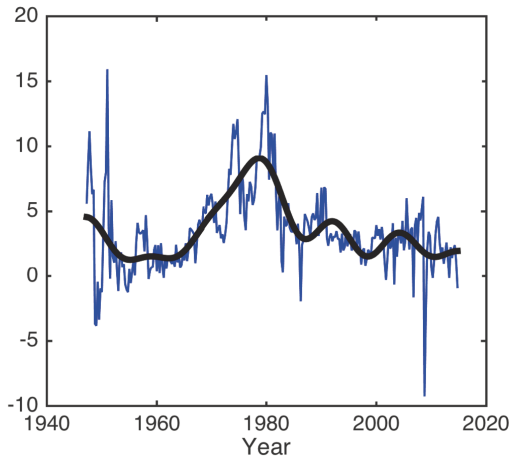


(ii) Cosine transform

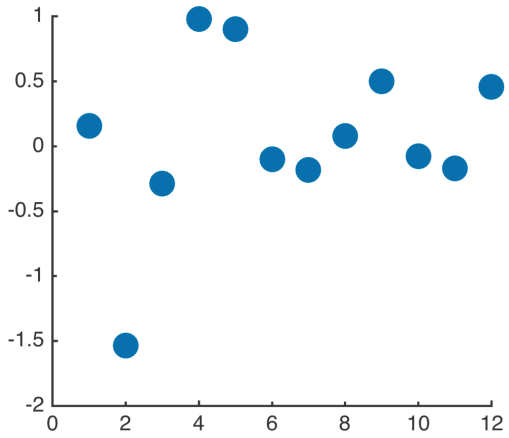


(b) Inflation (CPI)

(i) Series and low-frequency component



(ii) Cosine transform



Notes: The low-frequency components in (i) are the projection of the series onto $cos[(t\text{-}0.5)\pi j/T]$ for $j = 0, \ldots, 12$. The cosine transforms shown in (ii) are the self-normalized values $X^s_{T,1q}$.

the series onto $\cos[(t-1/2)\pi j/T]$ for $j = 0, ..., 12$. (The value $q = 12$ is used in the empirical analysis in Section 6 for reasons discussed below). The coefficients in the projection are the cosine transformations, $X_{T,1:q}$, and their standardized values, $X^s_{T,1:q}$ are plotted in panels (ii). These low-frequency components of the data are the summaries of the sample data we use to construct long-horizon prediction sets. Looking at panels (i), inflation exhibits much more low-frequency variation than GDP growth rates over the sample period; this is manifested in panels (ii) by the relatively larger magnitude of inflation's first few cosine transformations, capturing pronounced low-frequency movements. ▲

3 Statistical Preliminaries

The last section laid out the finite-sample prediction problem. In this section we review and develop the statistical theory that will guide our approach to constructing prediction sets. We divide the section into three subsections. The first provides a central limit theorem that characterizes the large-sample behavior of the weighted averages $(X_{T,1:q}, Y_T)$, and provides a characterization of the limiting covariance matrix based on the properties of the (pseudo-) spectrum of x_t near frequency zero. The second subsection illustrates this framework in the fractional $I(d)$ model and reports prediction sets for known d and Bayes prediction sets using a prior for d. The final subsection discusses the generic problem of robustifying Bayes prediction sets to obtain sets with frequentist coverage uniformly over the parameter space.

3.1 Large-Sample Approximations

To derive the asymptotic behavior of $(X_{T,1:q}, Y_T)$, note that each element can be written as a weighted average of x_t, $t = 1, \ldots, T + h$. 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

$$\eta_T = T^{1-\kappa} \int_0^{1+r} g(s) x_{\lfloor sT \rfloor + 1} ds \tag{4}$$

for a suitably chosen constant κ. In our context, the elements of $X_{T,1:q}$ are cosine transformations of the in-sample values of x_t (cf. (2)), 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 (3) 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 η_T.

In Appendix 8.1 we provide a central limit theorem for η_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 $\Delta x_t = \Delta x_{T,t}$ is a double array process with moving average representation $\Delta x_{T,t} = c_T(L)\varepsilon_t$, where ε_t is a possibly conditionally heteroskedastic martingale difference sequence with more than 2 unconditional moments, which allows for some forms of short memory stochastic volatility.[6] The moving average coefficients in $c_T(L)$ are square summable for each T, so that $\Delta x_{T,t}$ has a spectrum, denoted by $F_T(\lambda)$. The motivation for allowing $c_T(L)$ and F_T to depend on T is to capture many forms of persistence, as stemming from an autoregressive root local-to-unity, $\rho_T = 1 - c/T$, for instance.

The main regularity condition of the central limit theorem concerns the behavior of the (pseudo-) spectrum of $x_{T,t}$, $R_T(\lambda) = F_T(\lambda)/|1 - e^{-\mathrm{i}\lambda}|^2$, for frequencies close to zero. (In general, R_T is only a pseudo spectrum, since $\int_{-\pi}^{\pi} R_T(\lambda)d\lambda$ might not exist; for instance, it doesn't if $\Delta x_{T,t}$ is white noise, so that $x_{T,t}$ is a random walk). In particular, we assume that there exists a function $S : \mathbb{R} \mapsto \mathbb{R}$ such that for all fixed $K > 0$,

$$\lim_{T\to\infty} \int_0^K |T^{1-2\kappa} R_T(\frac{\omega}{T}) - S(\omega)|d\omega \to 0 \tag{5}$$

where S is such that $\int_0^\infty \omega^2 S(\omega)d\omega < \infty$. Intuitively, S describes the limiting behavior of R_T close to frequency zero, and we correspondingly denote it as the "local-to-zero" spectrum.

Under these and additional technical assumptions, Theorem 1 in the appendix shows that η_T has a limiting normal distribution,[7] and as an implication

$$T^{1-\kappa} \begin{bmatrix} X_{T,1:q} \\ Y_T \end{bmatrix} \Rightarrow \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(0, \Sigma) \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right) \tag{6}$$

with $X = (X_1, \ldots, X_q)'$ (we omit the dependence of X on q to ease notation). The asymptotic covariance matrix Σ is a function of the local-to-zero spectrum S, as discussed further below.

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 $Y^s_T =$

[6] The restriction $E[\Delta x_t] = 0$ rules out a deterministic trend in x_t. This restriction is plausible in our empirical analysis in which x_t denotes growth rates of real variables like per capita GDP, inflation rates, and asset returns.

[7] 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 η_T. This assumption might be violated by rare 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* $\bar{x}_{T+1:T+h}$ over a long horizon to invalidate a normal approximation.

$Y_T/\sqrt{X'_{T,1:q}X_{T,1:q}}$ follows directly from (6) and the continuous mapping theorem,

$$\begin{bmatrix} X^s_{T,1:q} \\ Y^s_T \end{bmatrix} \Rightarrow \begin{bmatrix} X^s \\ Y^s \end{bmatrix} = \begin{bmatrix} X/\sqrt{X'X} \\ Y/\sqrt{X'X} \end{bmatrix}. \tag{7}$$

Note that as a consequence of the imposed scale invariance, the convergence (7) holds irrespective of the scaling factor κ in (4), and the distribution of (X^s, Y^s) does not depend on the scale of Σ. Explicit expressions for the densities f_{X^s} and $f_{(X^s,Y^s)}$ of X^s and (X^s, Y^s) as a functions of Σ_{XX} and Σ are provided in Appendix 8.2.

With Σ known, it is straightforward to compute prediction sets of Y^s given $X^s = x^s$: A calculation shows that the distribution of Y^s conditional on $X^s = x^s$ satisfies (see Appendix 8.2)

$$\frac{Y^s - \Sigma_{YX}\Sigma_{XX}^{-1}x^s}{\sqrt{\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}\sqrt{x^{s\prime}\Sigma_{XX}^{-1}x^s/q}} \sim \text{Student-}t^q \tag{8}$$

so that prediction sets for Y^s of a given level $1-\alpha$ are readily computed using Student-t quantiles. These sets in turn imply asymptotically justified prediction sets for Y^s_T via (7), and thus also for $\overline{x}_{T+1:T+h}$ via the definition of $(X^s_{T,1:q}, Y^s_T)$ and (3).

In particular, when $x_{T,t}$ is $I(0)$ with long-run variance σ^2 (i.e., the local-to-zero spectrum is flat, $S(\omega) = (2\pi)^{-1}\sigma^2$), it turns out that $\Sigma_{YX} = 0$, $\Sigma_{XX} = \sigma^2 I_q$, $\Sigma_{YY} = (1 + r^{-1})\sigma^2$, and the $1-\alpha$ prediction set for Y is given by the interval with endpoints $\pm t^q_{(1-\alpha/2)}\sqrt{q(1+r^{-1})X'X}$, where $t^q_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of the student-t distribution with q degrees of freedom. The asymptotically justified prediction set for $\overline{x}_{T+1:T+h}$ is therefore $\overline{x}_{1:T} \pm t^q_{(1-a/2)}(1+r^{-1})^{1/2}T^{-1/2}s_{LR}$, where $s^2_{LR} = (T/q)X'_{T,1:q}X_{T,1:q}$. Note that this interval becomes *smaller* for a larger horizon r – a law of large numbers effect reduces the variability of the average of future values, with the residual uncertainty under $r \to \infty$ stemming from sampling uncertainty about the population mean $E[x_t]$.

More generally, the asymptotic covariance matrix Σ can always be expressed as a function of the local-to-zero spectrum S and the weighting functions g_j that correspond to the jth element of $(X', Y)'$. In particular, Corollary 1 of Appendix 8.1 implies that

$$\Sigma_{j,k} = \int_0^\infty S(\omega)w_{jk}(\omega)d\omega \tag{9}$$

where $w_{jk}(\omega) = 2\,\text{Re}[\left(\int_0^{1+r} g_j(s)e^{-\mathrm{i}\omega s}ds\right)\left(\int_0^{1+r} g_k(s)e^{\mathrm{i}\omega s}ds\right)]$. The elements of the covariance matrix of $(X', Y)'$ are thus weighted averages of the local-to-zero spectrum S, with weights $w_{jk}(\omega)$ that are functions of Fourier transform of the x_t weights $g_j(s)$ used to construct X and Y.

The weights $w_{jk}(\omega)$ are plotted in the supplementary appendix; we highlight three features here. First, $w_{jk}(\omega)$ with $j \neq k$, integrates to zero, which implies that for a flat local-to-zero spectrum S (corresponding to an $I(0)$ model), Σ is diagonal, as already noted above. Second, the weight associated with the predictor X_j is mostly concentrated in the interval $\pi j \pm 2\pi$, so the variance of X_j is determined by the value of S in this frequency band. Third, the weight associated with Y has its mass concentrated near $\omega = 0$; for example when $r = 1/2$, the variance of Y is mostly determined by the shape of S on the interval $\omega \in [0, 4\pi]$. The implication of these results is that the conditional variance of Y given X 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.

3.2 Prediction Sets in the $I(d)$ Model

A leading example of this analysis is given by the fractional $I(d)$ model, which has a (pseudo-) spectrum proportional to $|\lambda|^{-2d}$ for λ close to zero; this yields the local-to-zero spectrum $S(\omega) \propto |\omega|^{-2d}$, and the central limit result from the last subsection is applicable for $-1/2 < d < 3/2$. The $I(d)$ model captures a wide range of long-run dependence patterns including the usual $I(0)$ and $I(1)$ models, but also persistence patterns between and outside these two extremes. With negative values of d it also allows for long-run anti-persistence (which may arise from overdifferencing), and with $d > 1$ it allows for processes more persistent than an $I(1)$ process.[8]

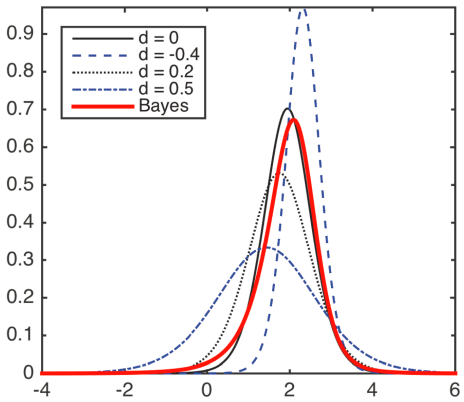
Running example (continued): Panels (i) of Figure 2 shows the appropriately centered and scaled Student-t predictive densities from (8) for the average growth rate of U.S. real per-capita GDP and the average value of CPI-inflation over the next 25 years for various values of d in the $I(d)$ model. For real GDP growth rates, predictive densities are shown for $d = -0.4$, 0.0, 0.2 and 0.5, and for inflation the predictive densities are shown for $d = 0.0$, 0.4, 0.7, and 1.0. For both series, as d increases, the variance of the predictive density increases because more persistence leads to larger variability in future average growth. The mode of the $I(0)$ predictive density is given by the in-sample mean (see the discussion following equation (8)), and the mode shifts to the left for $d > 0$ reflecting the persistent effect of the slow growth and low inflation experienced at the end of sample. In contrast, the mode of the $d = -0.4$ predictive density (shown for real GDP growth rates) is larger than the in-sample

[8] We discusss the numerical determination of Σ in the fractional model in the supplementary appendix.

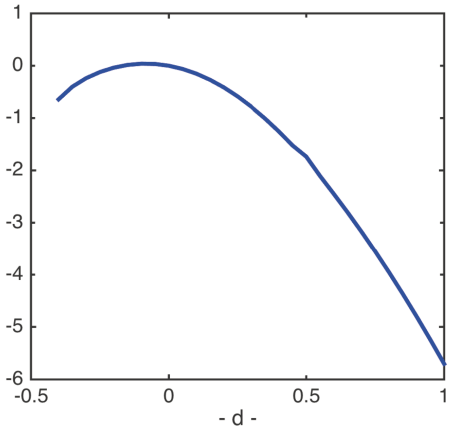
Figure 2: Predictive densities and low-frequency log likelihood values for the *I(d)* model

(a) Growth rate of real per-capita GDP

(i) 25-year ahead *I*(*d*)-predictive densities

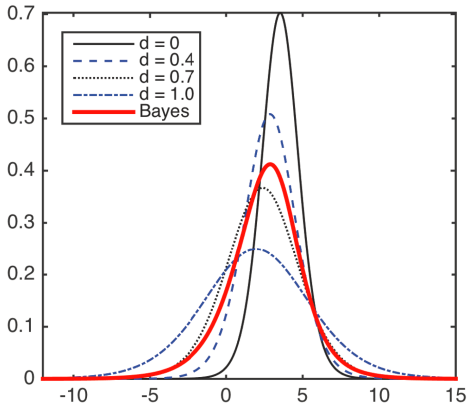


(ii) Low-frequency log-likelihood values for *d*

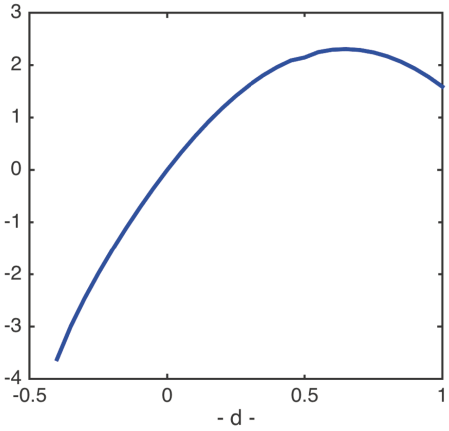


(b) Inflation (CPI)

(i) 25-year ahead *I*(*d*)-predictive densities



(ii) Low-frequency log-likelihood values for *d*



Notes: Panels (i) show the known-*d* prediction sets and Bayes prediction sets using the prior $d \sim$ U[-0.4,1.0]. The low-frequency *I*(*d*) likelihood is computed using $X^s_{T,1q}$ and its asymptotic distribution given in the text; values are relative to the *I*(0) model.

mean because faster than average growth is required to return the log-level of GDP to its pre-Great Recession trend growth path.

Evidently, both the length and location of 25-year ahead prediction sets depend critically on the d. This raises the question: What is the value of d for these series?

Panels (ii) summarize what the sample data say about the value of d. It plots the "low-frequency" log-likelihood values for d based on $X^s_{T,1:12}$ and its large-sample density from (7), and with the log-likelihood of the $I(0)$ model normalized to zero. The numbers for real per-capita GDP suggest only limited persistence 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 log-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. In contrast the inflation data suggest much more persistence: the likelihood has a maximum at $d = 0.65$ with corresponding log-likelihood value that is 2.3 points larger than the $I(0)$ model.

Taken together, the results in panels (i) and (ii) indicate that much of the 25-year-ahead forecast uncertainty is associated with uncertainty about the degree of persistence in the stochastic process, which in the $I(d)$ model is governed by the value of the parameter d. ▲

Bayes Prediction Sets: A natural way to incorporate this parameter uncertainty is to use a Bayes approach, where the limited sample information is combined with a prior on d. This is straightforward: With Γ the prior on d, the Bayes predictive density for Y^s conditional on $X^s = x^s$ is given by

$$f^{\Gamma}_{Y^s|X^s}(y^s|x^s) = \frac{\int f_{(X^s,Y^s)|d}(x^s, y^s)d\Gamma(d)}{\int f_{X^s|d}(x^s)d\Gamma(d)}$$

with $f_{(X^s,Y^s)|d}$ and $f_{X^s|d}$ the densities of (X^s, Y^s) and X^s in (7) with the value of Σ implied by a local-to-zero spectrum $S(\omega)$ proportional to $|\omega|^{-2d}$.

Bayes prediction sets can be readily computed from the predictive density. For example the "highest predictive density" (HPD) set for Y^s is $A^{HPD}(x^s) = \left\{y^s : f^{\Gamma}_{Y^s|X^s}(y^s|x^s) > \text{cv}(x^s)\right\}$, where $\text{cv}(x^s)$ solves $\int_{A^{HPD}(x^s)} f^{\Gamma}_{Y^s|X^s}(y^s|x^s)dy^s = 1 - \alpha$. This HPD Bayes set is the smallest length set that satisfies the coverage constraint relative to $f^{\Gamma}_{Y^s|X^s}$. Alternative Bayes prediction sets, such as equal-tailed sets, can be used instead. Thus, let $A^{\text{Bayes}}(x^s)$ denote a generic Bayes prediction set for Y^s as a function of x^s. Because $Y^s = Y/\sqrt{x'x}$ and $x^s = x/\sqrt{x'x}$, equivariance implies the extension to generic x via $A^{\text{Bayes}}(x) = \{y : y/\sqrt{x'x} \in A^{\text{Bayes}}(x/\sqrt{x'x})\}$.

Running example (continued): Panels (i) of Figure 2 shows the resulting Bayes predictive densities for $\bar{x}_{T:T+h}$ with a uniform prior on $d \in [-0.4, 1.0]$. This mixture of Student-t

densities is no longer necessarily symmetric, as the the underlying Student-t densities don't have the same mode. So for instance, for the GDP series, one obtains a left-skewed Bayes predictive distribution since larger values of d both increase uncertainty and shift the most likely future values to the left. ▲

3.3 Frequentist Robustification of Bayes Prediction Sets

As discussed in Section 3.1, the distributions of (X, Y) and (X^s, Y^s) depend on the covariance matrix Σ, which in turn depends on the low-frequency spectrum S of x_t. In the next section, we discuss a parameterization of the spectrum that is more general than the $I(d)$ model, so in general, $\Sigma = \Sigma(\theta)$ where θ is a parameter vector. In this section we discuss the general problem of constructing frequentist prediction sets that incorporate uncertainty about the value of θ. We provide additional details in Appendix 8.3.

The (frequentist) coverage probability of a set A, $P_\theta(Y \in A(X))$, generally depends on the value θ. A Bayes prediction set has coverage probability of $1-\alpha$, *on average* relative to the prior Γ, that is $\int P_\theta(Y \in A^{\text{Bayes}}(X))d\Gamma(\theta) = 1-\alpha$, but in general, $P_\theta(Y \in A^{\text{Bayes}}(X)) < 1-\alpha$ for some values of θ. In this subsection, we "robustify" Bayes sets by enlarging them so they have frequentist coverage: $\inf_{\theta\in\Theta} P_\theta(Y \in A(X)) \geq 1-\alpha$. There is no unique way to achieve this. We focus on sets with smallest weighted expected length.

To be specific, let $A(X)$ denote an arbitrary prediction set, and $V_\theta(A) = E_\theta[\text{vol}(A(X))]$ denote its expected length (which depends on θ). The goal is to choose A to minimize $V_\theta(A)$ over the parameter space Θ for θ. In many problems, including the one considered in this paper, there is no set A that simultaneously minimizes $V_\theta(A)$ for all $\theta \in \Theta$ while maintaining coverage, so there is an inherent trade-off of expected length over different values of θ. Let W denote a weighting function that makes this trade-off explicit. Consider the following problem:

$$\min_A \int V_\theta(A) dW(\theta) \tag{10}$$

subject to

$$\text{Equivariance:} \quad y \in A(x) \text{ implies } by \in A(bx) \text{ for all } x, y \text{ and } |b| \neq 0 \tag{11}$$

$$\text{Frequentist Coverage:} \quad \inf_{\theta\in\Theta} P_\theta(Y \in A(X)) \geq 1-\alpha, \text{ and} \tag{12}$$

$$\text{Bayes Superset:} \quad A^{\text{Bayes}}(x) \subset A(x) \text{ for all } x. \tag{13}$$

Because the objective function depends on the weighting function W, so will the solution, and we discuss specific choices for W in the following section. The constraint (11) imposes scale invariance – recall that location invariance in the original problem is imposed by the

choice of Y and X. The coverage constraint that defines a $(1-\alpha)$-frequentist prediction set is given by (12).

The constraint (13) can be motivated in a variety of ways. One motivation is *ad hoc* and simply says that the goal is to robustify a Bayes set by enlarging it so it has frequentist coverage properties. Another focuses on properties of frequentist sets that do not impose (13). Notably, conditional on particular realizations of X these sets can have unreasonably small length; indeed they can be empty. In particular, even with θ known (i.e., $\Theta = \{\theta\}$), solving (10) subject to (11) and (12) does *not* in general yield the known-θ prediction set (8), but rather a prediction set whose coverage of Y is equal to $1-\alpha$ only on average over repeated draws of X, but not conditional on the observed X. Müller and Norets (2012) show that imposing (13) eliminates these arguably unattractive properties. We find the Müller and Norets arguments compelling and therefore enforce the constraint (13) for the frequentist sets used in the empirical analysis of Section 6. However, for comparison we also study solutions that do not impose (13) in Section 4 and the supplementary appendix.

The solution to the program (10)-(13) can be found in three steps: the first step transforms the problem to impose equivariance (11); the second uses a "least favorable distribution" for θ to simplify the coverage constraint (12); and the third enforces (13). We discuss these steps in turn.

Equivariance: If $A(X)$ is scale equivariant, then $Y \in A(X)$ if and only if $Y \in \sqrt{X'X}A(X^s)$. Thus, $\text{vol}(A(X)) = \sqrt{X'X}\,\text{vol}(A(X^s))$ and $V_\theta(A) = E_\theta\left[g_\theta(X^s)\,\text{vol}(A(X^s))\right]$, where $g_\theta(X^s) = E_\theta[\sqrt{X'X}|X^s]$. Imposing this restriction, the objective function (10) becomes

$$\min_A \int E_\theta\left[g_\theta(X^s)\,\text{vol}(A(X^s))\right] dW(\theta), \tag{14}$$

and the coverage (12) and Bayes superset (13) constraints can be rewritten as

$$\inf_{\theta\in\Theta} P_\theta(Y^s \in A(X^s)) \geq 1-\alpha \tag{15}$$

$$A^{\text{Bayes}}(x^s) \subset A(x^s) \text{ for all } x^s. \tag{16}$$

Note that (14)-(16) only involve the value of A evaluated at x^s, which lives on a smaller subspace $x^{s\prime}x^s = 1$ compared to $x \in \mathbb{R}^q$, but on that subspace, A is unrestricted. The solution to (14) subject to (15) and (16), $A^*(x^s)$, then implies the solution $A^*(x) = \{y : y/\sqrt{x'x} \in A^*(x/\sqrt{x'x})\}$ to the original problem (10) subject to (11)-(13).

Frequentist Coverage: For the coverage constraint (15), suppose for a moment that θ is a random variable with distribution Λ, and consider solving (14) subject to the resulting single coverage constraint

$$\int P_\theta(Y^s \in A(X^s))d\Lambda(\theta) \geq 1-\alpha. \tag{17}$$

A calculations yields the solution

$$A_\Lambda(x^s) = \left\{ y^s : \frac{\int f_{(Y^s,X^s)|\theta}(y^s,x^s)d\Lambda(\theta)}{\int g_\theta(x^s) f_{X^s|\theta}(x^s)dW(\theta)} > \text{cv} \right\} \tag{18}$$

where cv is chosen to satisfy (17) with equality. Of course, while A_Λ satisfies the *average* coverage constraint (17), it does not necessarily satisfy the *uniform* coverage constraint (15) required for a frequentist prediction set. However, because any set satisfying (15) also satisfies (17), the value of the objective (14) evaluated at A_Λ provides a lower bound for any set satisfying (15). Therefore, *if* a distribution $\Lambda^\dagger$ can be found under which $A_{\Lambda^\dagger}$ satisfies (15), then $A_{\Lambda^\dagger}$ solves the minimization problem (14) subject to the uniform coverage constraint in (15). Such a $\Lambda^\dagger$ is called the "least favorable distribution" for the problem. Elliott, Müller, and Watson (2015) develop numerical methods for approximating least favorable distributions in related problems, and we use a variant of those methods here. See the supplementary appendix for details.

Bayes Superset: The final step – incorporating the constraint (16) – is straightforward: it simply amounts to replacing (18) with the set

$$A^{MN}(x^s) = \left\{ y^s : \frac{\int f_{(Y^s,X^s)|\theta}(y^s,x^s)d\Lambda^\dagger(\theta)}{\int g_\theta(x^s) f_{X^s|\theta}(x^s)dW(\theta)} > \text{cv}^{MN} \right\} \cup A^{\text{Bayes}} \tag{19}$$

where $(\Lambda^\dagger, \text{cv}^{MN})$ are now such that $\int P_\theta(Y^s \in A^{MN}(X^s))d\Lambda^\dagger(\theta) = 1-\alpha$ and $\inf_{\theta\in\Theta} P_\theta(Y^s \in A^{MN}(X^s)) \geq 1-\alpha$ (cf. Theorem 4 in Müller and Norets (2012)).

4 Parameterizations for Long-Horizon Prediction Sets

Implementation of the prediction sets discussed in the last section requires four ingredients: (i) a parameterization of S, the local-to-zero spectrum, which yields the covariance matrix $\Sigma(\theta)$ via (9); (ii) a Bayes prior $\Gamma(\theta)$, which yields the Bayes prediction set A^{Bayes}; (iii) a frequentist weighting function $W(\theta)$, which quantifies the trade-off of expected length for various of θ in the objective function (10); and (iv) a choice for q, the number of cosine weighted averages used for the prediction problem. These are discussed in the following three subsections.

4.1 Parameterizing the Low-Frequency Spectrum

The $I(d)$ model introduced in Section 3.2 above is a flexible one-parameter model that captures a wide range of long-run persistence patterns. Because of its simplicity, flexibility, and use in other empirical analyses involving long-run behavior of economic time series, we use the $I(d)$ parameterization for our equal-tailed Bayes prediction sets A^{Bayes}.

However, a concern is that the family of $I(d)$ local-to-zero spectra may not be sufficiently flexible to capture all forms of long-run dependencies in economic time series. This suggests the need for a richer class of local-to-zero spectra, and we construct such a class by considering two other models that have proven useful for modelling low-frequency characteristics in other contexts. The first is the local-level model which 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 $I(1)$ component has relative magnitude $1/b$ and is usefully thought of as a stochastically varying 'local mean' of the growth rate x_t, as arising from some forms of stochastic breaks. In this model, $S(\omega) \propto b^2 + \omega^{-2}$. The second model is the local-to-unity AR(1) model, widely used to model highly persistent processes. In this 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). The $I(d)$, local-level and local-to-unity models are nested in the parameterization

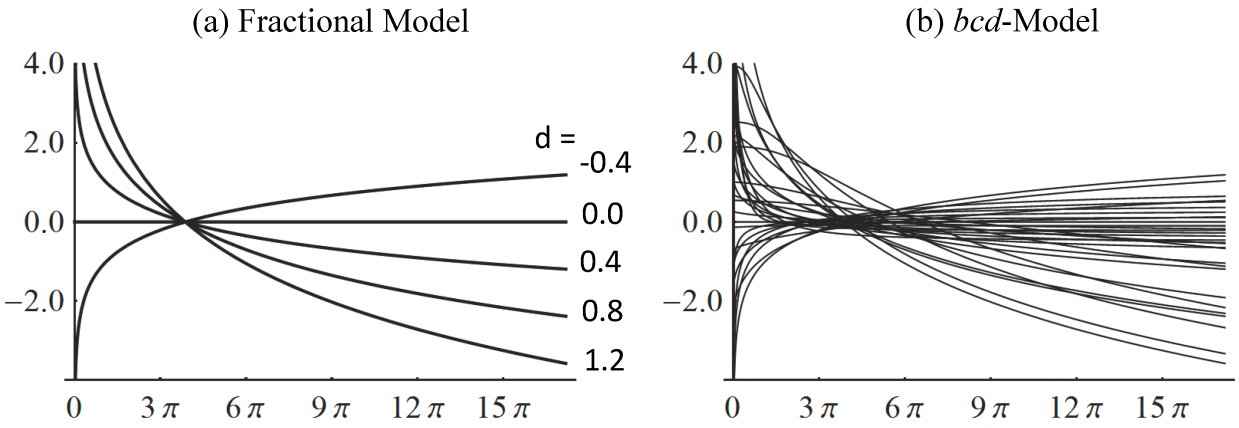
$$S(\omega) \propto \left(\frac{1}{\omega^2 + c^2}\right)^d + b^2 \tag{20}$$

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]

Figure 3 plots the logarithm of the local-to-zero spectrum of the $I(d)$ model in panel (a), and of this "*bcd*-model" in panel (b). The *bcd*-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. In the analysis below we let $\theta = (b, c, d)$, so that $\Sigma(\theta)$ is given by (9) with the local-to-zero spectrum S as in (20).

[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. It is also recognized as the spectrum of the Whittle-Matérn process from spatial statistics (e.g., Lindgren (2013)). Autocovariances for this process are derived in the supplementary appendix.

Figure 3: Logarithm of the local-to-zero spectrum for selected models



Note: The scale of the local-to-zero spectra S are normalized such that $\ln(S(\pi)) + \ln(S(15\pi)) = 0$.

4.2 Bayes and Frequentist Weighting Functions

In the empirical analysis in Section 6, we assume that S is characterized by the *bcd*-model, with $-0.4 \leq d \leq 1.0$ and $b, c \geq 0$.[10] As mentioned above, we construct Bayes sets using a prior that puts all weight on the $I(d)$ model (so that $b = c = 0$); we use a prior with uniform weight on values of $d \in [-0.4, 1.0]$. The A^{MN} sets robustify these Bayes sets so they have frequentist coverage for all values of $b, c \geq 0$ and $d \in [-0.4, 1.0]$. The analysis is usefully thought of in terms of the various spectral shapes plotted in Figure 3, and the Bayes prior is seen as putting equal weight on the various shapes in panel (a). Because S may take on shapes other than those represented by the $I(d)$ models in panel (a), the A^{MN} sets robustify the Bayes analysis to ensure frequentist coverage over all shapes shown in panel (b).

Construction of the A^{MN} sets requires specification of the weighting function W in (10). As noted in Section 3.3, the function W determines the trade-off between expected length for various of θ, which is necessary because there is no single prediction set that minimizes expected length for all θ. Our choice of W is guided by the observation that, even with θ known, the minimized values of $V_\theta(A)$ vary greatly over the values of θ. For example, in the $I(d)$ model with known d, prediction sets are much wider when $d = 1$ (so that $x_t \sim I(1)$) than when $d = 0$ ($x_t \sim I(0)$). To account for these differences we scale $V_\theta(A)$ so that it is expressed in units of the expected length of the predictions set for known θ. Denote this

[10] For the variables we study (growth rates of real variables, inflation rates, and asset returns), values of $d > 1.0$ are unnecessary, but these values may be appropriate in other applications, and we note that the results in Section 3.1 hold for the *bcd*-model with $-0.5 < d < 1.5$, and $b, c \geq 0$.

Table 1: Coverage for nominal 67% and 90% prediction sets, $r = 0.5$, $q = 12$

	67% Prediction Sets				90% Prediction Sets		
	A^{Bayes}	A_d^{MN}	$A_{(b,c,d)}^{MN}$		A^{Bayes}	A_d^{MN}	$A_{(b,c,d)}^{MN}$
$b = 0$, $c = 0$, $d \sim$U[-0.4,1.0]	0.67	0.71	0.73		0.90	0.93	0.94
Coverage minimized over:							
$b = 0$, $c = 0$, $-0.4 \leq d \leq 1.0$	0.55	0.67	0.68		0.81	0.90	0.90
$b \geq 0$, $c \geq 0$, $-0.4 \leq d \leq 1.0$	0.54	0.57	0.67		0.81	0.84	0.90

Notes: A^{Bayes} is the equal-tailed Bayes prediction set using the prior $b = 0$, $c = 0$, $d \sim$U[-0.4,1.0]; A_d^{MN} robustifies the Bayes set so it has frequentist coverage for $b = 0$, $c = 0$, $d \in$[-0.4,1.0] ; $A_{(b,c,d)}^{MN}$ robustifies the Bayes set so it has frequentist coverage for $b \geq 0$, $c \geq 0$, $d \in$[-0.4,1.0]

scaled version of $V_\theta(A)$ by $R_\theta(A) = V_\theta(A)/V_\theta^{\text{known}}$, where V_θ^{known} is the expected length of the prediction set for known value of θ implied by (8). In terms of $R_\theta(A)$ we use a weighting function that coincides with the Bayes prior: uniform on $d \in [-0.4, 1.0]$ and with $b = c = 0$ (so in terms of $V_\theta(A)$, the weighting function W is proportional to $1/V_\theta^{\text{known}}$).[11]

Table 1 shows coverage rates for 67% and 90% prediction sets for $h = rT$, with $r = 0.5$ using $q = 12$ cosine transforms. (This is the value of q we will use in the empirical analysis, and is discussed more fully in the next subsection). Table 1 answers two questions. First, what is the frequentist coverage of the Bayes prediction sets across the range of processes represented by the spectra in panels (a) and (b) of Figure 3? And second, does the $I(d)$ model provide sufficient flexibility so that the additional parameters b and c are unnecessary in practice? The table therefore displays coverage rates for three prediction sets: the Bayes set, A^{Bayes}; the set robustified to have correct frequentist coverage over d but with $b = c = 0$, denoted A_d^{MN}; and the set robustified to have correct frequentist coverage over (b, c, d), $A_{(b,c,d)}^{MN}$. Coverage rates are shown for three configurations of (b, c, d). In the first, values of (b, c, d) are drawn from the prior, so A^{Bayes} has correct coverage; in the second, the coverage probability is minimized over $-0.4 \leq d \leq 1.0$ with $b = c = 0$, so A_d^{MN} has correct coverage; and in the third, the coverage probability is also minimized over $b, c \geq 0$, so $A_{(b,c,d)}^{MN}$ has the correct coverage. The table indicates that A^{Bayes} exhibits substantial undercoverage for some values of d and (b, c, d). It also indicates substantial undercoverage of A_d^{MN} for some values of (b, c, d). Evidently, controlling coverage over d does not provide adequate coverage for long-run persistence patterns associated with non-zero values of b and c. Thus, because

[11] We investigate how this weighting function performs relative to other possible weighting in the supplementary appendix, where we also compute the cost (in terms of expected length) of the Bayes superset constraint (16).

some economic variables are arguably well-described by stochastic processes with non-zero value of b and c, it seems prudent to construct the $A^{MN}_{(b,c,d)}$ sets.[12]

We draw four conclusions from Table 1 and the analysis of W detailed in the supplementary appendix. First, Bayes sets constructed using a uniform prior on d exhibit substantial undercoverage for some values of d. Second, robustifying these sets to achieve frequentist coverage over d is inadequate for some processes with non-zero values of b and c. Third, for many values of (b, c, d) our benchmark choices of Γ and W produce sets with expected length close to the smallest achievable length under the coverage constraint. And finally, for most values of (b, c, d) there is little cost in terms of expected length for constructing frequentist sets that are supersets of Bayes sets (and therefore share some their desirable properties).

4.3 Choice of q

As discussed in Section 2, the choice of q may usefully be 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 3.1 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_t over the frequencies $[-q\pi/T, q\pi/T]$ is well approximated by the spectrum of the *bcd*-model becomes less plausible the larger q. Roughly speaking, we fit a parametric model to the q observations $X_{T,1:q}$, so a concern about nontrivial approximation errors arises for large q, irrespective of the sample size T.

We are thus faced with a classic efficiency and robustness trade-off. Recall from the discussion in Section 3.1, that the object of interest – the variability of long-run forecasts, as embodied by the conditional variance of Y given X – is a low frequency quantity that is essentially governed by properties of x_t over frequencies $[-12\pi/T, 12\pi/T]$. Since the

[12] The approximate least favorable distributions (ALFDs) that underlie the $A^{MN}_{(b,c,d)}$ sets are plotted in the supplementary appendix. The ALFD is non-degenerate and has most of its mass on spectra that are relatively flat for larger ω, but with a pronounced pole at zero (these spectra arise, for instance, in the local-level with moderate b). Intuitively, in the local-level model, the strong mean reversion of the $I(0)$ component masks the pronounced long-run uncertainty, making it relatively hardest to control coverage.

Table 2: 25-year-ahead prediction sets

Series	67%			90%		
	A^{Bayes}	A^{MN}	$A^{I(0)}$	A^{Bayes}	A^{MN}	$A^{I(0)}$
GDP/Pop	(1.3, 2.6)	(1.2, 2.6)	(1.4, 2.5)	(0.7, 3.0)	(-0.3, 3.0)	(0.9, 3.0)
Inflation	(0.3, 4.8)	(0.3, 4.8)	(2.3, 4.8)	(-1.8, 6.5)	(-2.6, 6.8)	(1.4, 5.7)

Notes: The table shows the prediction sets for the average value of the growth rate of real per-capita GDP and inflation from 2014-2039 using the benchmark values of Γ, W, and q.

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 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 $2T^{-1/2}t^q_{(1-a/2)}(1 + r^{-1})^{1/2}E\sqrt{X'X/q}$ with $X \sim \mathcal{N}(0, \sigma^2 I_q)$, which decreases in q, since $t^q_{(1-a/2)}E\sqrt{X'X/q}$ is a decreasing function of q.[13] 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 A.2 in the supplementary appendix quantifies the combined scale and shape effects by reporting the value of the objective $\int V_\theta(A)dW(\theta)$ in the program (10) for $q \in \{6, 12, 24, 48\}$. In this θ unknown case, there is an 8% decrease in average length as q increases from $q = 12$ to $q = 24$ and a further reduction of 5% for $q = 48$.

In our view, these potential gains are still relatively moderate and do not outweigh concerns about spectral misspecification that arise with a large choice of q. We therefore suggest constructing the prediction sets by default with $q = 12$, but also report results for different values of q in Section 6 below.

Running example (continued): Table 2 shows the 67% and 90% A^{Bayes} and A^{MN} 25-year-ahead predictions sets for real GDP growth and inflation using the benchmark values of the Bayes prior (Γ), weighting function (W), and $q = 12$. The 67% A^{Bayes} and A^{MN} sets

[13] This is analogous to the wider confidence intervals that arise from the use of inconsistent HAC estimators as developed by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2005), for example; see Müller (2014) for a review.

coincide, while the 90% A^{MN} sets are somewhat wider than the A^{Bayes} sets. For comparison, the table also shows the prediction sets computed from the $I(0)$ model. These are similar to the A^{Bayes} and A^{MN} sets for GDP (although the 67% $I(0)$ set is shifted to the left for reasons discussed above), but are much different for inflation (where the $I(0)$ are shifted the right and are much narrower), and where both results are as expected given the predictive densities and log-likelihood values displayed in Figure 2. Section 6 discusses these empirical results in more detail. ▲

5 Finite Sample Experiments

In the last two sections we developed a large-sample framework for constructing Bayes and frequentist long-run prediction sets that is tailored to models of long-run persistence typically used for economic time series. This large sample analysis is sufficiently general to allow for in-sample and out-of-sample stochastic breaks in the series, as long as these breaks occur with sufficient frequency that sample averages satisfy the central limit theorem discussed in Section 3. And the large-sample analysis also accommodates short memory stochastic shifts in volatility. But does this large-sample analysis provide reliable prediction sets for the sample sizes and stochastic processes typically encountered in applied economics? This section addresses this question using two sets of finite sample experiments. The first set of experiments are Monte Carlo simulations in which we generate data with level and volatility breaks designed to mimic the kinds of breaks seen in some macroeconomic time series. The second set of experiments uses rolling samples of daily interest rates and stock returns to construct pseudo-out-of-sample prediction sets and uses actual values of returns to evaluate these sets. We discuss these experiments in the following two subsections.

5.1 Monte Carlo Simulations with Breaks in Level and Volatility

Post-sample breaks of arbitrary size can undermine *any* attempt at prediction, so the methods proposed here are not immune to arbitrarily defined breaks. That said, a more relevant concern is how well the methods fare in the face of breaks that plausibly *have* occurred in the kinds of series to which the methods are to be applied. We address that question in this subsection. Statistical characterizations of uncertainty require a probability framework, so we consider breaks that occur probabilistically. And, because of the macroeconomic applications we carry out in Section 6, the models for these breaks are motivated by the behavior of important macroeconomic time economic series in the post-WWII United States.

We consider five models. The first two involve breaks in the level of x_t

$$x_t = \mu_t + u_t \tag{21}$$

where μ_t denotes the "level" of x_t and u_t is a zero-mean stochastic process that is independent of μ_t. We suppose that μ_t shifts discretely by an amount $\pm\delta$ at irregular time periods determined by the indicator s_t, so that

$$\mu_t = \mu_{t-1} + s_t\delta_t \tag{22}$$

where s_t is an i.i.d. Bernoulli process with $P(s_t = 1) = p$, and $\delta_t = \pm\delta$ with equal probability independent of s_t. Because μ_t follows a martingale, an $I(1)$ process, its sample averages are characterized by the Gaussian limits in Section 3 (as an $I(1)$ model for fixed $p, \delta > 0$ and a special case of the local-level model in subsection 4.1 for fixed $p > 0$ and $\delta = O(T^{-1})$). That said, when p is small, shifts in μ_t occur infrequently and the finite sample behavior of sample averages may be quite different from their large-sample Gaussian limit.

The second two models involve breaks in volatility. In these models x_t has components that can be represented as $\sigma_t e_t$, where e_t is an $I(0)$ process and σ_t is a volatility process that evolves as $\ln(\sigma_t) = \mu_t$, where μ_t follows (22). While the central limit used in Section 3 allows for certain forms of heteroskedasticity, it does not allow volatility to evolve as an $I(1)$ process. Thus, the volatility models in this section involve stochastic processes that are strictly more general than the processes analyzed above, even in large samples.

The final model involves breaks in both the level and volatility of x_t. Specifically, following Pesaran, Pettenuzzo, and Timmermann (2006) (also see Chib (1998)), x_t follows a different AR(1) process (with intercept and possibly a unit root) within each a sequence of regimes. Regimes end stochastically according the Bernoulli process s_t above, although with a regime-specific value of p, and new regimes begin with new parameter values for the AR process and p drawn from a fixed probability distribution.

We choose model parameters to match specific characteristics of post-WWII U.S. quarterly macroeconomic data. Thus, we chose T to correspond to 65 years, and as above we consider forecast horizons of $h = 0.5T$ with $q = 12$ and the prior (Γ) and weighting function (W) described in Section 4. For models 1-4, we choose two values for the break frequency: $p_{\text{large}} = 1/40$ (so a break occurs, on average, once every 40 quarters) and $p_{\text{small}} = 1/260$ (so a break occurs, on average, once during the sample period period). The other parameter values depend on the experiment and are motivated by the behavior of particular U.S. macroeconomic time series.

Model 1 is motivated by the growth rate of average labor productivity, which visually appears to be an $I(0)$ process but around a time varying level. (See the supplementary

appendix Figure B.4.) Labor productivity growth averaged 2.2% per year in the post-WWII period, but experienced decade-long swings that were roughly one percentage point higher (early 1960s and late 1990s) or lower (1970s and early 1980s) than the average. The first model therefore takes the form (21) with $u_t \sim iid\mathcal{N}(0, \sigma_u^2)$, where σ_u is chosen to match the long-run standard deviation of average labor productivity, and the magnitude of the breaks in μ_t was chosen to yield a sensible value for the interquartile range (IQR) of $\mu_T - \mu_0$. Specifically, for each value of p we chose two values for δ, where the first, δ_{small}, yielded an IQR of 0.5% and the second, δ_{large}, yielded an IQR of 1.5%.

Model 2 is similar to Model 1, but is motivated by the behavior of nominal interest rates, which follow a pattern consistent with (21) but with u_t a highly serially correlated process. Thus for this experiment, u_t was generated by an AR(1) process with coefficient 0.98, Gaussian innovations with variance chosen to match 10-year U.S. Treasury Bonds, and δ_{small} and δ_{large} chosen so that the IQR for $\mu_T - \mu_0$ was 2.0% and 4.0%, respectively.

Model 3 is designed to capture features in the data like the "Great Moderation": a low-frequency reduction in the volatility in real U.S. macroeconomic variables. For example, the standard deviation of growth rates of measures of real aggregate activity (GDP, employment, etc.) fell rather abruptly by roughly 30% in the early 1980s (e.g., Stock and Watson (2002)). Thus, in this model the data were generated as $x_t = \sigma_t e_t$, with $e_t \sim iid\mathcal{N}(0, 1)$ and $\ln(\sigma_t) = \mu_t$ generated as described above with δ_{small} and δ_{large} chosen so that the IQR for $\ln(\sigma_T/\sigma_0)$ was 0.25% and 0.75%, respectively.

Model 4 is designed to capture the changes in variability and persistence evident in the U.S. inflation process. Stock and Watson (2007), Cogley and Sargent (2014), and others argue that these features can be captured by a local-level-model with stochastic volatility. Thus, in this model we generate data as $x_t = e_{1t} + \sum_{s=1}^{t} \sigma_s e_{2s}$ where e_{1t} and e_{2t} are mutually independent i.i.d. standard normal random variables, $\ln(\sigma_t) = \mu_t$ follows the process described above, and the parameters are chosen to mimic estimates of the time-varying $IMA(1,1)$ representation of the model found in U.S. data (e.g., Watson (2014)). Specifically, σ_0 is chosen so that MA coefficient is 0.5 in the initial period, and δ_{small} and δ_{large} were chosen so that the IQR of the full-sample change in the MA coefficient was 0.5 and 0.8.

Model 5 uses parameter values from Pesaran, Pettenuzzo, and Timmermann (2006) of their analysis of nominal U.S. Treasury Bill rates from 1947-2002 and therefore captures the changing volatility and persistence in post-WWII interest rates.[14]

Results for the various experiments are shown in Table 3, where panel (a) shows results for the A^{Bayes} sets and panel (b) shows results for the A^{MN} sets. The first row of each panel shows results for the model with $p = 0$ (so that breaks are absent); the other rows show

[14]The specific parameter values are discussed in the supplementary appendix.

Table 3: Coverage probability for simulated data, T = 65 years, r = 0.5, and q = 12

		Model 1: Break in level, $I(0)$ model			Model 2: Break in level, AR(1) model			Model 3: Break in volatility, $I(0)$ model			Model 4: Break in volatility, $I(0)$+$I(1)$ model			Model 5: PPT Break Model	
Nom. Level		67%	90%		67%	90%		67%	90%		67%	90%		67%	90%
(a) A^{Bayes}															
$p = 0$		0.72	0.94		0.65	0.89		0.72	0.94		0.55	0.81		0.65	0.88
p_{large}	δ_{small}	0.68	0.92		0.64	0.89		0.71	0.94		0.57	0.77			
	δ_{large}	0.58	0.85		0.63	0.88		0.71	0.92		0.58	0.73			
p_{small}	δ_{small}	0.69	0.93		0.64	0.89		0.71	0.94		0.57	0.79		0.62	0.80
	δ_{large}	0.62	0.88		0.63	0.88		0.71	0.93		0.57	0.75			
(b) A^{MN}															
$p = 0$		0.74	0.95		0.70	0.94		0.74	0.95		0.67	0.90		0.72	0.93
p_{large}	δ_{small}	0.71	0.94		0.71	0.94		0.74	0.95		0.67	0.85			
	δ_{large}	0.67	0.90		0.70	0.93		0.73	0.94		0.65	0.79			
p_{small}	δ_{small}	0.72	0.94		0.71	0.93		0.74	0.95		0.68	0.87		0.68	0.85
	δ_{large}	0.69	0.91		0.71	0.93		0.74	0.94		0.66	0.82			

Notes: The table shows coverage probability of A^{Bayes} and A^{MN} sets for four models subject to breaks in level (Models 1 and 2) or volatility (Models 3 and 4), or level and volatility (Model 5). In Models 1-4 breaks occur in each time period with probability p and are of size δ_{small} or δ_{large}. The models are described in the text. Models 1-4 use 65 years of quarterly data. Model 5 uses 65 years of monthly data.

results for p_{small}, p_{large}, δ_{small} and δ_{large}. When $p = 0$, Models 1 and 3 are i.i.d. processes for which both A^{Bayes} and A^{MN} have coverage rates that exceed their nominal level. This overcoverage occurs because A^{Bayes} provides correct *average* coverage for $I(d)$ processes that includes both small and large values of d, and coverage for small d is less than the average coverage. Similar reasoning explains the overcoverage for A^{MN}, which is designed to achieve uniform coverage over (b, c, d). And with $p = 0$, Model 2 is well approximated by the local-to-unity model with $c = 260(1 - 0.98) = 5.2$ and Model 4 is well approximated by an $I(1)$ process; A^{MN} satisfies the coverage constraint in both models, while A^{Bayes} severely undercovers in model 4, achieving the same undercoverage shown previously in Table 1 for the $I(d)$ model. Moving to the results with $p > 0$, A^{Bayes} has coverage rates notably less than its nominal level in Models 4 and 5; coverage rates for nominal 67% A^{MN} are approximately correct for all models, but there is some undercoverage in Models 4 and 5.

In summary, we conclude that the build-in safeguards against non-stationarities in our approach seem to be mostly adequate for series that are comparable to post-WWII U.S. macroeconomic series.

5.2 Pseudo-out-of-sample Forecasts

The last section examined the performance of long-run prediction sets using simulated data, but how well do the sets perform for actual data? Ideally, pseudo-out-of-sample experiments could be used to answer this question using economic time series from a wide array of stochastic processes. However, this is difficult in our setting – where we are interested in long-horizon forecasts for macroeconomic series in developed economies like the U.S. – because the available macroeconomic data provide little pseudo-out-sample information.

But the salient definition of a long-run forecast is that the horizon is long relative to the sample data. And in contrast to macroeconomic data, there are long time series on high-frequency financial variables. One empirical test of the methods developed here is thus to see whether forecasts constructed from, say, one year of financial data, have reasonable empirical coverage for forecasts of the average value over the following half year. We carry out two pseudo-out-of sample experiments.

For the first experiment we use value-weighted S&P daily returns from CRSP from 1926-2014, for a total of $23,535$ returns. The pseudo-out-of-sample exercise uses a rolling sample of $T = 260$ observations to construct prediction sets for the average value of r_t and r_t^2 over the next $h = 0.5T = 130$ periods, where the choice of T matches the sample size used in the last section and in much of the empirical analysis in Section 6. Rolling through the sample in this way allows us to compute $23,145$ (or 178 non-overlapping) pseudo-out-sample

Table 4: Coverage rates for prediction sets in pseudo-out-of-sample experiments
Rolling sample, $T = 260$ days, q=12 and $r = 0.5$

Prediction Set	Returns			Squared returns			3-Month Treasury Bill Interest Rate	
	67%	90%		67%	90%		67%	90%
A^{Bayes}	0.71	0.92		0.64	0.84		0.53	0.73
A^{MN}	0.73	0.93		0.69	0.88		0.65	0.85
$A^{I(0)}$	0.67	0.88		0.42	0.69		0.11	0.24

Notes: The table shows empirical coverage rates for A^{Bayes} and A^{MN} prediction sets in 23,145 (=178 non-overlapping) pseudo-out-of-sample periods for the average value of SP500 returns and squared returns, and 15,007 (=115 non-overlapping) periods for the average value of 3-Month Treasury Bill interest rate.

prediction sets. The second experiment is similar, but uses daily observations on nominal interest rates for 3-month U.S. Treasury Bills from 1954-2014.

Results for 67% and 90% prediction sets are summarized in Table 4. For the return series, both A^{Bayes} and A^{MN} have sample coverage rates slightly larger than their nominal values; this result is not unexpected given the results in the preceding sections. Squared returns are significantly more persistent than the level of returns, and are often given as an example of an economic time series that exhibits $I(d)$ low-frequency behavior (see, for instance, Ding, Granger, and Engle (1993)). Table 4 indicates that the pseudo-out-of-sample coverage for A^{Bayes} is slightly lower than its nominal level, while the coverage of A^{MN} remains near its nominal level; again, these results are not unexpected given the simulation results of the last subsection. Daily values of nominal interest rates are highly persistent and exhibit shifting volatility; coverage rates for A^{Bayes} are substantially below their nominal levels, while coverage rates for A^{MN} are much closer to the nominal level; these results are broadly in line with those from Model 5 of the last section. In contrast, for squared returns and the interest rate series, forecast intervals computed from the $I(0)$ model have coverage far below nominal level, underlying the necessity to flexibly adjust to various forms of persistence.

5.3 A Final Pseudo-out-of-sample Forecast

The results from the Monte Carlo simulations lead us to conclude that predictions sets based on asymptotic approximations developed in Sections 3 and parameterizations in Section 4 perform reasonably well in the face of the kinds of breaks that have occurred in the post-WWII U.S. macroeconomy. This conclusion is buttressed by the results from the pseudo-out-of-sample forecasts for daily asset returns and interest rates. Of course, this does not imply that these prediction sets will produce sensible *ex-post* results in all circumstances, and we end this section with one example.

Data on per-capita GDP suggest that the U.S. economy was dramatically more volatile in the pre-WWII period than after. For example, the standard deviation of annual per-capita GDP growth rates fell from 7.8% over 1901-1946 to just 2.4% over 1947-2014. Estimates of long-run standard deviations show a similar reduction (8.5% falling to 2.6%). While the source of the decline is a matter of debate (see Balke and Gordon (1989), Romer (1989), and Watson (1994) for discussion), imagine using the data from 1901-46 to construct a prediction set for average growth over the following 46 years, from 1947-1992. Using the formula below equation (8), the $I(0)$ prediction set is $\bar{x}_{1901:1946} \pm t^{12}_{1-\alpha} \times 2 \times 46^{-1/2} \times s_{LR}$, where s_{LR} is the estimated long-run standard deviation constructed from the pre-war data with $q = 12$. Using $\bar{x}_{1901:1946} = 1.86\%$ and $s_{LR} = 8.54\%$, the 67% prediction set is wide: $(-0.7\%, 4.4\%)$. Indeed, *given* the low volatility experienced since 1947, the prediction set is implausibly wide; had it been constructed using the post-1946 value of $s_{LR} = 2.6\%$ it would have been much narrower, $(1.1\%, 2.6\%)$. (The realized value of average GDP growth over 1947-1992 was $\bar{x}_{1947:1992} = 2.1\%$.)

What do we make of the 1946 prediction set? Here are two observations. First, there *was* considerable uncertainty about the future of U.S. growth following WWII, with many forecasters predicting a return to the growth patterns experienced during the 1930s and others predicting rapid growth (see Walton and Rockoff (2013)). The 1946 prediction set was arguably more plausible in 1946 than it is today. Second, while the Monte Carlo simulations suggested relatively small coverage distortions associated with low-frequency volatility shifts, these shifts (i) were not as large as the 2.5-fold decrease in volatility in post-WWII GDP and (ii) were two-sided (volatility increases and decreases), while the single realization for GDP was necessarily one-sided. A lesson from this example is that in some circumstances it may be important to explicitly incorporate large and potentially predictable breaks in volatility, and the required modifications are outlined in the paper's final section.

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 using sample data through 2014. These series include the growth rate of per-capita values of real GDP and CPI inflation used as the running examples, and also the growth rates of real per-capita consumption expenditures, population, productivity (both total factor and labor productivity), real stock returns, and prices as measured by the PCE deflator. 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.

Table 5: Prediction sets

a. 67% coverage

Series	Forecast horizon (in years)			
	10	25	50	75
Quarterly Post-WWII Series				
GDP/Pop	(1.1, 3.0)	(1.3, 2.6) [1.2, 2.6]	(1.5, 2.4) [0.7, 2.4]	(1.5, 2.4) [0.5, 2.4]
Cons/Pop	(1.2, 3.0)	(1.4, 2.7)	(1.5, 2.6) [1.2, 2.6]	(1.6, 2.5) [1.0, 2.5]
TF Prod	(0.3, 1.8) [0.1, 1.8]	(0.5, 1.8) [-0.2, 1.9]	(0.6, 1.8) [-0.4, 2.2]	(0.6, 1.8) [-0.6, 2.4]
Labor Prod	(0.8, 2.6) [0.8, 2.7]	(1.0, 2.6)	(1.2, 2.5) [0.8, 2.7]	(1.3, 2.5) [0.6, 2.8]
Population	(0.6, 1.0) [0.5, 1.0]	(0.5, 1.1) [0.4, 1.2]	(0.4, 1.2) [0.2, 1.4]	(0.4, 1.3) [0.0, 1.5]
Inflation (PCE)	(0.2, 3.7)	(0.0, 4.1)	(-0.2, 4.5)	(-0.4, 4.7)
Inflation (CPI)	(0.4, 4.5)	(0.3, 4.8)	(0.2, 5.1)	(0.1, 5.3) [-0.1, 5.3]
Infl. (CPI,Japan)	(-1.4, 4.4) [-1.7, 4.4]	(-1.7, 4.9) [-2.4, 4.9]	(-2.1, 5.4) [-3.8, 5.4]	(-2.4, 5.7) [-4.7, 6.3]
Stock Returns	(1.8 15.3)	(2.6, 13.8)	(3.0, 13.1) [2.9, 13.1]	(3.1, 12.9) [1.3, 13.0]
Longer Span Data Series				
GDP/Pop	(0.2, 4.5)	(0.9, 3.4)	(1.2, 2.9)	(1.4, 2.7)
Cons/Pop	(0.2, 2.6)	(0.6, 2.5) [0.6, 2.6]	(0.8, 2.4) [0.5, 2.9]	(0.8, 2.4) [0.3, 3.1]
Population	(0.5, 1.2)	(0.5, 1.3)	(0.5, 1.4)	(0.5, 1.4) [0.4, 1.4]
Inflation (CPI)	(-0.2, 5.9)	(0.3, 5.6)	(0.6, 5.4)	(0.7, 5.4)
Stock Returns	(0.7, 13.2)	(2.9, 11.0)	(3.8, 9.9)	(4.2, 9.5)

Table 5: Prediction sets
(continued)

b. 90% coverage

Series	Forecast horizon (in years)			
	10	25	50	75
Quarterly Post-WWII Series				
GDP/Pop	(0.3, 3.7)	(0.7, 3.0) [-0.3, 3.0]	(0.8, 2.8) [-0.6, 2.9]	(0.9, 2.7) [-0.9,3.2]
Cons/Pop	(0.4, 3.7)	(0.7, 3.1) [0.0, 3.1]	(0.8, 3.0) [-0.3, 3.1]	(0.9, 2.9) [-0.6, 3.4]
TF Prod	(-0.3, 2.4) [-0.6, 2.4]	(-0.2, 2.3) [-1.1, 2.7]	(-0.1, 2.3) [-1.5, 3.1]	(-0.1, 2.4) [-1.9, 3.5]
Labor Prod	(0.0, 3.2) [0.0, 3.4]	(0.0, 3.0) [-0.2, 3.1]	(0.0, 3.1) [-0.6, 3.6]	(0.0, 3.2) [-0.9, 3.9]
Population	(0.4, 1.2) [0.3, 1.2]	(0.3, 1.3) [0.1, 1.5]	(0.1, 1.5) [-0.2, 1.7]	(0.0, 1.6) [-0.5, 2.0]
Inflation (PCE)	(-1.1, 5.0) [-1.3, 5.0]	(-1.8, 5.6) [-2.7 6.1]	(-2.6, 6.5) [-4.4, 7.8]	(-3.2, 7.1) [-5.8, 9.2]
Inflation (CPI)	(-1.2, 6.0) [-1.4, 6.0]	(-1.8, 6.5) [-2.6, 6.8]	(-2.5, 7.2) [-4.4, 8.5]	(-3.1, 7.9) [-5.8, 9.8]
Infl. (CPI,Japan)	(-3.7, 6.7) [-4.7, 6.7]	(-4.7, 7.7) [-6.8, 8.8]	(-5.9, 9.0) [-9.6, 11.7]	(-6.9, 10.0) [-12.0, 14.0]
Stock Returns	(-3.1, 21.0)	(-2.0, 20.1) [-3.8, 20.1]	(-2.2, 20.2) [-7.3, 22.7]	(-2.5 20.5) [-9.8, 25.2]
Longer Span Data Series				
GDP/Pop	(-1.4, 6.2) [-1.8, 6.5]	(-0.1, 4.3)	(0.4, 3.6)	(0.7, 3.3)
Cons/Pop	(-0.7, 3.5) [-0.7, 4.0]	(-0.3, 3.1) [-0.6, 3.5]	(-0.2, 3.0) [-0.9, 3.9]	(-0.1, 3.0) [-1.3, 4.2]
Population	(0.3, 1.5)	(0.2, 1.6) [0.1, 1.6]	(0.1, 1.7) [-0.2, 1.8]	(0.0, 1.8) [-0.4, 2.0]
Inflation (CPI)	(-2.6, 8.2)	(-2.3, 7.8)	(-2.4, 8.0) [-2.9, 8.1]	(-2.7, 8.3) [-3.7, 9.0]
Stock Returns	(-3.9, 18.4) [-5.7, 18.4]	(-0.2, 15.2)	(1.1, 13.9)	(1.6, 13.4) [1.2, 13.4]

Notes: This table shows the 67% and 90% prediction sets for forecast horizons, $h = 10, 25, 50$, and 75 years. The A^{Bayes} sets are shown in parentheses and are based on the $I(d)$ model with uniform prior for $-0.4 \leq d \leq 1.0$. The A^{MN} sets are shown in brackets, and are omitted if they coincide with the A^{Bayes} sets. By construction the A^{MN} sets control asymptotic coverage in the *bcd*-model with $-0.4 \leq d \leq 1.0$, and b and c unrestricted.

Sources and details of construction of the data are presented in the supplementary Data Appendix. Supplementary appendix Figures B.1-B.14 provide a variety of summary statistics for each series including a plot of the series, its low-frequency components, normalized cosine transformations, low-frequency $I(d)$ log-likelihood values, and 67% and 90% Bayes, MN and $I(0)$ prediction sets for all horizons between 10 and 75 years. Table 5 reports a summary of the prediction sets for prediction sets for 10, 25, 50, and 75 year horizons.

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

Real per capita GDP. The Bayes prediction 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. The frequentist sets coincide with the Bayes sets for (relatively) short horizons but include smaller values of average GDP growth rates at longer horizons. Apparently, to guarantee high coverage uniformly in the *bcd*-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 (1901-2014) 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 discussed above.

At the 75-year horizon the 80% Bayes prediction interval (not shown) is 1.3 to 2.5, which roughly coincides with the 80% 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 μ) 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 log-likelihood values plotted in Figure 2 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 essentially coincides with the Bayes set.

Productivity. The log-likelihood values for d indicate that productivity (TFP and average labor productivity) may have somewhat greater than $I(0)$ persistence; see Figures B.3 and B.4. This translates into prediction sets that are wider than $I(0)$ sets, particularly for frequentist sets at large forecast horizons. 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. U.S. population growth shows considerable low-frequency variability over the 20th century and the post-WWII period. Immigration and fertility dynamics are presumably at the source of these long swings. The low-frequency MLE of d is very close to unity over both sample periods, with the $I(1)$ log-likelihood more than 7 points higher than in the $I(0)$ model. Table 5 and Figures B.5 and B.12 show prediction intervals that widen as the forecast horizon increases, a natural characteristic of $I(1)$ predictive densities. 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 both Bayes and frequentist sets. For example, while the 10-year 67% Bayes prediction set for U.S. CPI inflation is $(0.4, 4.5)$, the 90% set widens to $(-1.2, 6.0)$.

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 2013. (See Figure B.8.) Moreover, they are in line with predictive densities derived from asset prices. For example, Kitsul and Wright (forthcoming) 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%.[15] The corresponding probabilities computed from the Bayes predictive density constructed using the post-WWII data are 11% for deflation and 28% for high inflation.

Stock Returns. 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

[15] See Kitsul and Wright (forthcoming), Figures 3 and 4. Fleckenstein, Longstaff, and Lustig (2013) estimate somewhat lower probabilites for deflation, but similar probabilities for inflation exceeeding 4%. (See their Figures 4 and 5). For a related calculation, see Figure 3 in Hilsher, Raviv, and Reis (2014).

by the $I(0)$ model. For example, at the 25-year horizon, the 67%-$I(0)$ prediction set (from Figure B.9) is $(3.4, 11.2)$ while the corresponding Bayes and MN prediction sets (from Table 7) are $(1.8, 15.3)$. The longer-span data suggest somewhat less persistence ($\widehat{d}^{MLE} = -0.2$ for the 1926-2014 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. 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. Consistent with this, we find Bayes prediction sets that narrow as h increases, but more slowly than $h^{-1/2}$.

Results for different values of q. As discussed in Section 4, 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 6, which shows the 67% and 90% prediction sets for the 25-year ahead forecasts for $q = 6, 12$, and 24. 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 67% MN prediction set is 11% narrower using $q = 24$ than with $q = 12$ broadly consistent the results discussed in Section 4.

7 Conclusions

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. As we show, these averages conveniently have an approximate joint normal distribution under fairly general conditions. However, the prediction problem

Table 6: Prediction sets for various values of q, 25-year horizon

	67 %				90 %		
	$q = 6$	$q = 12$	$q = 24$		$q = 6$	$q = 12$	$q = 24$
Quarterly Post-WWII Data							
GDP/Pop	(0.6, 2.3) [0.2, 2.3]	(1.3, 2.6) [1.2, 2.6]	(1.5, 2.6)		(-0.1, 2.7) [-0.8, 2.7]	(0.7, 3.0) [-0.3, 3.0]	(0.9, 3.0) [0.0, 3.0]
Cons/Pop	(0.9, 2.6)	(1.4, 2.7)	(1.4, 2.6)		(0.0, 3.1) [-0.5, 3.1]	(0.7, 3.1) [0.0, 3.1]	(0.7, 3.1) [0.0, 3.1]
TF Prod	(0.1, 1.6) [-0.2, 1.6]	(0.5, 1.8) [-0.2, 1.9]	(0.6, 1.8) [-0.2, 2.0]		(-0.5, 2.1) [-1.1, 2.3]	(-0.2, 2.3) [-1.1, 2.7]	(0.1, 2.2) [-1.0, 2.7]
Labor Prod	(1.0, 2.6)	(1.0, 2.6)	(1.2, 2.5) [0.9, 2.5]		(0.2, 3.1) [-0.4, 3.1]	(0.0, 3.0) [-0.2, 3.1]	(0.5, 3.0) [-0.2, 3.2]
Population	(0.5, 1.2) [0.4, 1.3]	(0.5, 1.1) [0.4, 1.2]	(0.5, 1.1) [0.4, 1.1]		(0.3, 1.5) [0.0, 1.6]	(0.3, 1.3) [0.1, 1.5]	(0.3, 1.3) [0.1, 1.4]
Inflation (PCE)	(0.4, 4.7)	(0.0, 4.1)	(-0.2, 3.8)		(-1.7, 6.3) [-3.1, 6.5]	(-1.8, 5.6) [-2.7 6.1]	(-2.0, 5.3) [-2.6, 5.4]
Inflation (CPI)	(0.6, 5.3)	(0.3, 4.8)	(0.0, 4.5)		(-1.8, 7.0) [-3.3, 7.2]	(-1.8, 6.5) [-2.6, 6.8]	(-2.1, 6.2) [-2.5, 6.2]
Infl. (CPI,Japan)	(-1.6, 5.2) [-3.7, 5.2]	(-1.7, 4.9) [-2.4, 4.9]	(-1.3, 4.9)		(-4.7 8.0) [-8.2, 9.7]	(-4.7, 7.7) [-6.8, 8.8]	(-3.9, 7.7) [-5.5, 7.9]
Stock Returns	(2.0, 13.1)	(2.6, 13.8)	(2.7, 12.8)		(-3.1, 18.8)	(-2.0, 20.1) [-3.8, 20.1]	(-0.6, 16.4) [-9.0, 24.0]
Longer Span Data Series							
GDP/Pop	(0.8, 2.9)	(0.9, 3.4)	(0.8, 3.2)		(-0.2, 3.8)	(-0.1, 4.3)	(-0.3, 4.1)
Cons/Pop	(0.7, 2.5)	(0.6, 2.5) [0.6, 2.6]	(0.9, 2.6) [0.9, 2.7]		(-0.1, 3.3) [-0.3, 3.4]	(-0.3, 3.1) [-0.6, 3.5]	(0.2, 3.3) [-0.2, 3.7]
Population	(0.8, 1.6)	(0.5, 1.3)	(0.5, 1.3)		(0.4, 1.9) [0.3, 2.0]	(0.2, 1.6) [0.1, 1.6]	(0.2, 1.5) [0.1, 1.5]
Inflation (CPI)	(0.3, 6.1)	(0.3, 5.6)	(0.8, 5.3)		(-2.3, 8.7)	(-2.3, 7.8)	(-1.1, 7.0) [-1.6, 7.5]
Stock Returns	(2.0,11.5)	(2.9, 11.0)	(2.8, 9.2)		(-2.4, 15.8)	(-0.2, 15.2)	(0.4, 11.8) [-0.4, 11.8]

Notes: A^{Bayes} sets are shown in parentheses and A^{MN} sets are shown in brackets when they differ from Bayes sets. See notes to Table 5 for additional information.

is complicated because the covariance matrix of the limiting normal distribution depends on the shape of the spectrum over very low frequencies, and there is limited sample information about this 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 then 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 robustify the Bayes sets by enlarging them so that, by construction, they control coverage uniformly over all values of the *bcd*-parameters. We construct minimum expected length 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 narrower than 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 probability statements about (very) long-run forecasts.

While the analysis presented in this paper accommodates a wide range of low-frequency persistence patterns, it was not designed to directly accommodate large breaks in volatility such as those evident in the pre- and post-WWII U.S. GDP growth rate data. In principle it is possible to explicitly account for non-negligible non-stationarities in the volatility process by postulating a stochastic process for the volatility path, and integrating out this additional source of uncertainty (similar to the approach of Müller and Watson (2008) in their Section 3.3, say).

Also, our analysis has been univariate in the sense that we have constructed predictions sets for a scalar random variable $\overline{x}_{T+1:T+h}$ using sample values of x_t. However, answers to some questions require multivariate prediction sets. The statistical theory discussed and developed in Section 3 carries over directly to multivariate settings. That said, there are important practical obstacles to constructing multivariate prediction sets. These obstacles

include finding a convenient, but flexible, parameterization of the multivariate local-to-zero spectrum, constructing accurate approximations to least favorable distributions with high dimensional θ, and computing accurate approximations to the density of relevant invariants. Overcoming these obstacles is left to future research.

References

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

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.

Chib, S. (1998): "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 75, 221–241.

Cogley, T., and T. Sargent (2014): "Measuring Price Level Uncertainty and Stability in the U.S., 1850-2012," *forthcoming in the Review of Economics and Statistics*.

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.

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

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 (2015): "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

Fleckenstein, M., F. Longstaff, and H. Lustig (2013): "Deflation Risk," *NBER working paper w19238*.

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

Hilsher, J., A. Raviv, and R. Reis (2014): "Inflation Away the Public Debt? An Empirical Assessment," *NBER Working Paper w 20339*.

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 (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.

Kitsul, Y., and J. H. Wright (forthcoming): "The Economics of Options-Implied Inflation Probability Density Functions," *Journal of Financial Economics*.

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

Lindgren, G. (2013): *Stationary Stochastic Processes: Theory and Applications*. Chapman & Hall/CRC.

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

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

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.

——— (2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81.

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

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

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

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.

Romer, C. (1989): "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869-1908," *Journal of Political Economy*, 97(1), 1–37.

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

Stock, J. H. (1996): "VAR, Error Correction and Pretest Forecasts at Long Horizons," *Oxford Bulletin of Economics and Statistics*, 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 (2002): "Has the Business Cycle Changed and Why?," in *NBER Macroeconomics Annual 2002*, ed. by M. Gertler, and K. S. Rogoff, pp. 159–218. MIT Press, Cambridge, MA.

——— (2007): "Why Has Inflation Become Harder to Forecast?," *Journal of Money, Credit, and Banking*, 39, 3–34.

WALTON, G., AND H. ROCKOFF (2013): *History of the American Economy*. South-Western College Publishing, 12th edition edn.

WATSON, M. (1994): "Business cycle durations and postwar stabilization of the US economy," *American Economic Review*, 84(1), 24–46.

WATSON, M. W. (2014): "Inflation Persistence, the NAIRU, and the Great Recession," *American Economic Review*, 104, 31–36.

8 Appendix

8.1 Central Limit Theorem of 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*

$$E(\varepsilon_t^2 - 1|\mathcal{F}_{t-m}) \leq \xi_m \tag{23}$$

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) 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,

$$\int_0^K |F_T(\frac{\omega}{T}) - \omega^2 S(\omega)| d\omega \to 0; \tag{24}$$

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

$$T^{-3} \int_{K_T/T}^{\pi} F_T(\lambda)\lambda^{-4} d\lambda = \int_{K_T}^{\pi T} F_T(\omega/T)\omega^{-4} d\omega \to 0; \tag{25}$$

(iii.c)

$$T^{-3/2} \int_{1/T}^{\pi} F_T(\lambda)^{1/2}\lambda^{-2} d\lambda = T^{-1/2} \int_1^{\pi T} F_T(\omega/T)^{1/2}\omega^{-2} d\omega \to 0; \tag{26}$$

(iv) for some fixed integer H, *the function* $g : [0, H] \mapsto \mathbb{R}$ *is of bounded variation and satisfies* $\int_0^H g(s)ds = 0$.

Then

$$T^{-1/2} \int_0^H g(s) x_{T,\lfloor sT \rfloor + 1} ds \Rightarrow \mathcal{N}(0, \int_{-\infty}^{\infty} S(\omega) \left| \int_0^H e^{-\mathrm{i}\omega s} g(s) ds \right|^2 d\omega) \tag{27}$$

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 ε_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 (23).

The linear coefficients $c_{T,s}$ are scaled by the sample size T such that the convergence (27) holds with the same scaling factor $T^{-1/2}$ across various types of persistence, such as $I(0)$ and $I(1)$ models. See below for examples. Given our interest in scale equivariant prediction sets, this scale normalization is without loss of generality.

Since for any fixed K, $\sup_{0\leq\omega\leq K}|T^{-2}\frac{\omega^2}{|1-e^{-\mathrm{i}\omega/T}|^2}-1|\to 0$, assumption (iii.a) is equivalent to (5) (with $\kappa=3/2$).

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 $c_{T,s}$, $|c_{T,s}|\leq C_0e^{-C_1s}$ 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 $L>0$, $\omega^2S(\omega)$ is constant and equal to $(2\pi)^{-1}$ times the long-run variance of $\Delta x_{T,t}$, and (25) and (26) 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}, $c_{T,s}\approx C_0T^{-d}s^{d-1}$, 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 λ close to zero, $F_T(\lambda)\approx(2\pi)^{-1}C_0^2(\lambda T)^{-2d}$, so that $\omega^2S(\omega)=(2\pi)^{-1}C_0^2\omega^{-2d}$, and (25) and (26) 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, $F_T(\lambda)=(2\pi)^{-1}|1-e^{-\mathrm{i}\lambda}|^2/|1-\rho_Te^{-\mathrm{i}\lambda}|^2$, which is seen to satisfy (24) with $S(\omega)=(2\pi)^{-1}(\omega^2+c^2)^{-1}$. Conditions (25) and (26) 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 ε_0, with a scaling such that $F_T(\lambda)$ is $O_p(1)$ for $\lambda=O(T^{-1})$). Here $F_T(\lambda)=(2\pi)^{-1}T^2|1-e^{-\mathrm{i}\lambda}|^2=(2\pi)^{-1}4T^2\sin(\lambda/2)^2$, so that $S(\omega)=(2\pi)^{-1}$, and (25) evaluates to $4(2\pi)^{-1}\int_{K_T}^{\pi T}T^2\sin(\omega/2T)^2\omega^{-4}d\omega\leq(2\pi)^{-1}\int_{K_T}^{\pi T}\omega^{-2}d\omega\to 0$, and (25) to $2(2\pi)^{-1/2}T^{-1/2}\int_1^{\pi T}T\sin(\omega/2T)\omega^{-2}d\omega\leq(2\pi)^{-1/2}T^{-1/2}\int_1^{\pi T}\omega^{-1}d\omega\to 0$, where the inequalities follow from $\sin(\lambda)\leq\lambda$ for all $\lambda\geq 0$.

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

Corollary 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 $g_j(s)=\mathbf{1}[s\leq 1]\sqrt{2}\cos(\pi js)$ for $j=1,\ldots,q$. 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 **??** and the Cramer-Wold device via

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

since $g(s) = \sum_{j=1}^{q+1} \lambda_j g_j(s)$ clearly satisfies the assumption in Theorem 1 (iv). ■

8.2 Density of (X^s, Y^s) and Related Results

Let $Z = (X', Y)'$ and $U = \sqrt{X'X}$. Write μ_l for Lebesgue measure on $\mathbb{R}^l$, and ν_q for the surface measure of a q dimensional unit sphere. For $x \in \mathbb{R}^q$, let $x = x^s u$, 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 $y = y^s u$ with $u \in \mathbb{R}^+$ and $y^s \in \mathbb{R}$, so that $d\mu_1(y) = u d\mu_1(y^s)$. 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

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

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

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

where the second equality follows from the form of the Gamma density function, and Γ denotes the gamma function. The implied marginal density of X^s is

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

Similarly, with $g(x^s) = E[\sqrt{X'X} | X^s = x^s]$, we obtain

$$\begin{aligned} f_{x^s}(x^s) g(x^s) &= \int_0^\infty u f_{(X^s, U)}(x^s, u) d\mu_1(u) \\ &= u (2\pi)^{-q/2} |\Sigma_{XX}|^{-1/2} \int_0^\infty u^{q-1} \exp[-\tfrac{1}{2} u^2 (x^{s\prime} \Sigma_{XX}^{-1} x^s)] d\mu_1(u) \\ &= 2^{-1/2} \pi^{-q/2} |\Sigma_{XX}|^{-1/2} \Gamma(\tfrac{q+1}{2}) (x^{s\prime} \Sigma_{XX}^{-1} x^s)^{-(q+1)/2}. \end{aligned}$$

Finally, from (6), $\tilde{Y} = Y - \Sigma_{YX}\Sigma_{XX}^{-1}X \sim \mathcal{N}(0, \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY})$ and X are independent normal random variables. Also, using well known properties of a multivariate standard normal distribution, $X'\Sigma_{XX}^{-1}X \sim \chi_q^2$ is independent of $\tilde{X}^s = \Sigma_{XX}^{-1/2}X/\sqrt{X'\Sigma_{XX}^{-1}X}$. Since X^s is a one-to-one transformation of $\tilde{X}^s$, we thus obtain

$$\frac{\tilde{Y}}{\sqrt{X'\Sigma_{XX}^{-1}X/q}\sqrt{\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}}|X^s \sim \text{Student-}t^q$$

and the result (8) follows by dividing the numerator and denominator by $\sqrt{X'X}$.

8.3 Approximate Least Favorable Distributions

In practice, it won't be possible to compute a least favorable distribution $\Lambda^\dagger$ that perfectly solves the program (14)-(16). To make further progress, we follow Elliott, Müller, and Watson (2015) (EMW in the following), and first formally state a lower bound on (14), and then define an approximate least favorable distribution (ALFD) Λ^* that solves (10) within a tolerance of ϵ.

To ease notation, write $V_W(A) = \int V_\theta(A)dW(\theta)$ and $C_\theta(A) = P_\theta(Y^s \in A(X^s))$. Also, we make the dependence of the set (18) on cv explicit by writing

$$A_{\Lambda,\text{cv}}(x^s) = \left\{ y^s : \frac{\int f_{(Y^s,X^s)|\theta}(y^s, x^s)d\Lambda(\theta)}{\int g_\theta(x^s)f_{X^s|\theta}(x^s)dW(\theta)} > \text{cv} \right\}. \tag{28}$$

We begin by proving the optimality of the set $A_{\Lambda,\text{cv}}$ in the problem $\min_A V_W(A)$ subject to $\int C_\theta(A)d\Lambda(\theta) = 1 - \alpha$.

Lemma 1 *Let $A_{\Lambda,\text{cv}}$ be such that $\int C_\theta(A_{\Lambda,\text{cv}})d\Lambda(\theta) = 1-\alpha$. Then $A_{\Lambda,\text{cv}}$ solves $\min_A V_W(A)$ subject to $\int C_\theta(A)d\Lambda(\theta) \geq 1 - \alpha$.*

Proof. 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, $V_W(A) = \int\int\int g_\theta(x^s)f_{X^s|\theta}(x^s)\varphi(y^s, x^s)d\nu_q(x^s)d\mu_1(y^s)dW(\theta) = \int \varphi(z^s)f_1(z^s)d\lambda_{q,1}(z^s)$, and $\int C_\theta(A)d\Lambda(\theta) = \int\int\int f_{Z^s|\theta}(x^s, y^s)\varphi(y^s, x^s)d\nu_q(x^s)d\mu_1(y^s)d\Lambda(\theta) = \int \varphi(z^s)f_0(z^s)d\lambda_{q,1}(z^s)$, where $d\lambda_{q,1}(z^s) = d\nu_q(x^s) \times d\mu_1(y^s)$, $f_1(z^s) = \int g_\theta(x^s)f_{X^s|\theta}(x^s)dW(\theta)$ and $f_0(z^s) = \int f_{Z^s|\theta}(z^s)d\Lambda(\theta)$. Thus, the problem is equivalent to the problem of finding the best test that rejects (that is $\varphi = 1$) with probability at least $1 - \alpha$ when the "density" of Z^s is f_0, and minimizes the rejection probability when the "density" of Z^s is f_1. These densities do not necessarily integrate to one, but the solution still has to be of the Neyman-Pearson form (18), as can be seen by the very argument that proves the Neyman-Pearson Lemma: Set $\varphi^*(y^s, x^s) = \mathbf{1}[y^s \in A_{\Lambda,\text{cv}}(x^s)]$ and $\varphi(y^s, x^s) = \mathbf{1}[y^s \in A(x^s)]$ for some A that satisfies $\int C_\theta(A)d\Lambda(\theta) \geq 1 - \alpha$. Then $\int \varphi f_0 d\lambda_{q,1} \geq 1 - \alpha$ (we drop z^s as the dummy variable of integration for notational convenience), and

$$0 \quad \leq \quad \int (\varphi^* - \varphi)(f_0 - \text{cv}\, f_1)d\lambda_{q,1}$$

$$\leq \quad \text{cv}(\int \varphi f_1 d\lambda_{q,1} - \int \varphi^* f_1 d\lambda_{q,1})$$

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

A second result mirrors Lemma 1 of EMW and bounds the value of $\min_A V_W(A)$, formalizing the result verbally stated in Section 3.3.

Lemma 2 *Let $A_{\Lambda,\text{cv}}$ as in Lemma 1. Then for any A that satisfies $\inf_\theta C_\theta(A) \geq 1-\alpha$, $V_W(A) \geq V_W(A_{\Lambda,\text{cv}})$.*

Proof. The result is immediate from Lemma 1 after noting that $\inf_\theta C_\theta(A) \geq 1-\alpha$ implies $\int C_\theta(A) d\Lambda(\theta) \geq 1-\alpha$. ■

Lemma 2 is useful, as it provides a set of lower bounds (indexed by Λ) on the achievable values of the objective (14). Thus, if a Λ can be identified that implies a small lower bound in the sense that a small adjustment to the critical value yields a set with uniform coverage and only marginally larger objective, the problem has been solved as a practical matter. Again following EMW, we denote such a distribution an ALFD.

Definition 2 *An ϵ-approximate least favorable distribution Λ^* is a probability distribution on θ satisfying*

(i) there exists cv^* *such that $A_{\Lambda^*,\text{cv}^*}$ satisfies $\int C_\theta(A_{\Lambda^*,\text{cv}^*}) d\Lambda^*(\theta) = 1-\alpha$*

(ii) there exists $\text{cv}^{*\epsilon} < \text{cv}^*$ *such that* $\inf_\theta \int C_\theta(A_{\Lambda^*,\text{cv}^{*\epsilon}}) \geq 1-\alpha$, *and* $V_W(A_{\Lambda^*,\text{cv}^{*\epsilon}}) \leq V_W(A_{\Lambda^*,\text{cv}^*}) + \epsilon$.*

The strategy is thus to set some small tolerance level ϵ, and to numerically identify an ϵ-ALFD Λ^*. By definition, $A_{\Lambda^*,\text{cv}^{*\epsilon}}$ controls coverage uniformly, and invoking Lemma 2, its W-weighted average length is at most ϵ larger than of any prediction set that controls coverage uniformly.

Generalizations of Lemmas 1 and 2 for $A(x^s)$ additionally restricted to be a superset of some given set $B(x^s)$ are proven entirely analogously and are omitted for brevity (cf. Lemma 3 in EMW and Theorem 4 in Müller and Norets (2012)).