How to catch a star? Reliability of filtering estimates in linear state space systems

Gianni Amisano Thorsten Drautzburg Andrea Stella*

March 25, 2026

Preliminary and incomplete. Please do not circulate

Abstract

In macroeconomic policy analysis, estimates of latent variables such as the natural rate of interest and the output gap are important inputs for policymakers. We develop reliability measures to assess the quality of estimates of latent variables in linear state space models that go beyond describing the level of uncertainty surrounding these estimates. Motivated by the (infeasible) population regression of the true state at time $t + h$ on time t information and the time $t + h$ forecast, we propose the (conditional) correlation and the associated R^2 as intuitive measures to assess the predictability of the latent states. We relate this measure to the posterior variance of the state estimate. To help understand the source of the posterior uncertainty we propose the half-life and the contribution of the initial state to the conditional forecast error variance as additional diagnostics. Models that estimate the level of the hidden state very imprecisely may still imply predictability at relevant horizons. The Holston-Laubach-Williams (HLW) model features high state uncertainty and low predictability for the natural rate of interest. However, the relative uncertainty surrounding the output gap is only half as large in the HLW model, and reliability is about three times higher for the output gap. More generally, uncertainty and reliability are distinct concepts, which we illustrate with a simple model of potential output, trend inflation, and trend unemployment. Since it is impossible to directly assess whether the models accurately quantify the uncertainty and reliability of the state estimates, we propose tests whether the model quantify the uncertainty and reliability of observables correctly. For the HLW model, we find that the empirical model performance is largely consistent with the theoretical prediction.

*Amisano and Stella: Federal Reserve Board of Governors. Drautzburg: Federal Reserve Bank of Philadelphia. We would like to thank Jonathan Adams, Frank Diebold, Domenico Giannone, Frank Schorfheide, Jonathan Wright and several seminar and conference participants for comments. Any remaining errors are our own. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, or its Board of Governors.

1 Introduction

Modern macroeconomic policy fundamentally depends on accurately estimating unobservable economic variables. Designing effective stabilization policy requires knowing whether the economy operates above or below its potential, as errors in these estimates can lead to significant economic costs such as heightened inflation or unnecessary unemployment. Similarly, calibrating optimal monetary policy necessitates precise knowledge of the natural rate of interest, a critical input for central banks aiming to maintain economic stability. Consequently, government agencies, private companies, and academic researchers allocate substantial resources to estimating these hidden time series and debating the efficacy of various econometric models.

While it has been well-known that model estimates of these latent variables are often fraught with uncertainty[1], practitioners regularly use these models while paying relatively little attention to the accuracy or "goodness of fit" of these estimates. Because these quantities are inherently unobservable, traditional model validation techniques such as in-sample or out-of-sample fit metrics are ineffective or can only provide indirect indications.

We do not dismiss regression models solely because they yield wide standard errors. Likewise, we argue that analyses based on state-space models should be complemented by an evaluation of their ability to recover latent states. In particular, we propose quasi-analytical expressions for R^2 statistics in stationary and non-stationary models to assess how effectively a linear state-space model recovers the relevant unobservable time series, commonly termed state variables. A crucial first step in evaluation involves determining whether a model successfully identifies and recovers the associated shocks. Invertibility (e.g., Sims and Zha (2006) and Fernández-Villaverde et al. (2007)) and recoverability (e.g., Chahrour and Jurado (2022), Buncic, Pagan and Robinson (2023)) are well-known concepts in macroeconometrics. If a state or a shock is invertible, it can be recovered from an infinite history of present and past data. This corresponds to the Kalman filter estimate, or nowcast. Conversely, recoverability extends this notion to include future observations, allowing a state or shock to be inferred from the entire trajectory of past, present, and future data. This corresponds to the Kalman smoother, or backcast. Our proposed R^2 measure is in line with previous literature (see Sims and Zha (2006), Forni, Gambetti and Sala (2019), Plagborg-Møller and Wolf (2022)). A low R^2 value implies that the filtered estimate explains little of the variation in the truth, questioning the usefulness of the state space model.

What values of R^2 are considered acceptable is application specific. However, we show that there is a one-to-one mapping between our R^2 measure and the smoothness of the HP-filter (Hodrick and Prescott, 1997), when applied to the state-space representation of the HP filter. Macroeconomists typically calibrate this parameter, implying priors over R^2 that depend on the frequency of the model. Leveraging our quasi-analytical characterization of R^2, we show how researchers with hypotheses about R^2 can test them in

[1]Laubach and Williams (2003) report in Table 1 an average standard error of 1.9 for r_t^* and of 3.0 for $y*$ in their baseline specification, for example.

frequentist frameworks or use them as a prior in Bayesian frameworks. In non-stationary models such as the HP-filter, the unconditional R^2 is not defined. Instead, we introduce a conditional R^2 that assesses reliability at some horizon $h \geqslant 1$. This measure applies both in stationary and non-stationary settings.

Both estimation uncertainty and R^2-measures are functions of the parameters of the state space model. All else equal, there is a tight relationship between these measures: If the ex-post state uncertainty rises, the R^2 falls. However, different sources of uncertainty – signal noise or persistence or variability of the state itself – can break this relationship. Indeed, we provide a simple example where both state uncertainty and the R^2 are increasing in the persistence of the state. To better understand the source of uncertainty, we propose the half-life of state estimation errors and the contribution of the unknown initial state to the variation in the state as complementary measures.[2]

It is important to distinguish stationary and non-stationary models or variables. In line with Engle and Yoo (1987), we show that in non-stationary settings the state estimate is cointegrated with the true state. However, second moments of non-stationary variables do not exist and we therefore propose to work with conditional variances. We then need to take a stand on the horizon h of interest and our R^2 then quantifies how much the data flow from t to $t + h$ helps us reduce the accumulated uncertainty over the same period. Due to cointegration, this measure converges to 1 in non-stationary models as h diverges, but not in stationary models.

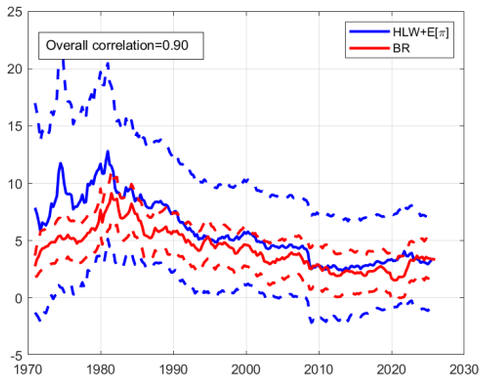
State uncertainty and our measures depend on the data only via the parameter estimates. It is therefore important to ask whether the model and the parameter estimates adequately quantify the uncertainty and reliability in the data. Since we cannot assess the empirical performance of the model-implied uncertainty and reliability measures for the latent states, we propose to do so for the observables. As one test, we examine whether the empirical and the model-implied RMSEs are consistent over the horizons of interest. In addition, we assess the reliability measures for the observables at difference horizons, and examine whether the gap between the model-implied reliability for observables is significantly different from the empirical counterparts. This is particularly important in models with unit roots, in case these mask potential mean-reversion at longer horizons and may lead to mistakenly inflate the implied R^2. For the HLW model, we find that the model reliability measures in the model are in line with those in the data at relevant horizons. However, the model overpredicts the inflation RMSE at horizons longer than two years.

Employing our proposed evaluation tools, we examine two prominent models frequently used to estimate the natural rate of interest (r^*, or its nominal counterpart i^*): those by Bauer and Rudebusch (2020) and Holston, Laubach and Williams (2023). As shown in Figure 1, their estimates of i^* are different and come with different levels of uncertainty. Our analysis demonstrates that the Bauer and Rudebusch (2020) model accurately recovers the natural rate of interest, achieving rapid convergence. Interestingly, this model relies exclusively on government bond yields, neglecting informative signals from prices, employment, and output markets. In contrast, the more comprehensive macroeconomic framework developed by Holston, Laubach and Williams (2023), although inclusive of broader economic indicators, consistently underperforms in

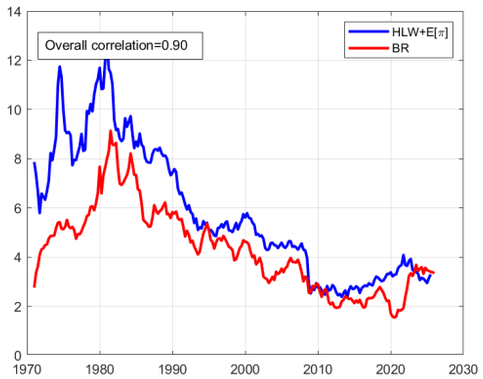
[2] All three measures are tied together by the steady state Riccati equation.

Figure 1: Filtered estimates of $r_t^* + E[\pi]$ (LW) and i_t^* (BR)

With uncertainty



Point estimates only



accurately recovering $i^* = r^* + E[\pi]$, aligning with findings from prior research.

We explore whether using more information in the HLW model improves its reliability. First, we explore whether a prior that the model's r^* estimates are reliable improves the model fit. However, we reject this restriction. Second, we augment the HLW with return expectations as a noisy measure of r^*, in the spirit of Kim and Wright (2005). This extension yields higher reliability measures and lower uncertainty.

The rest of the paper is structured as follows. Section 2 describes the econometric framework adopted by this paper, i.e. the standard linear state-space framework. Section 3 and 4 develop the reliability measures. Section 5 presents the empirical applications. Section 6 discusses assessing the predictive performance of the underlying state space model.

Literature. There is an established literature in macroeconometrics using R^2 statistics to measure whether shocks are revealed by the data. Fernández-Villaverde et al. (2007) develop a condition under which all shocks in a DSGE model are invertible in a VAR, i.e., can be recovered from the current and past VAR observables. Sims and Zha (2006) and Forni, Gambetti and Sala (2019) propose $1-R^2$ of a regression of the true shocks on the VAR counterparts to quantify the failure of invertibility. Chahrour and Jurado (2022) extend the notion of invertibility to recoverability – whether shocks can be recovered from past, present, and future data. Plagborg-Møller and Wolf (2022) use an R^2 statistics to quantify invertibility and recoverability of shocks. Similarly, we use R^2 statistics to quantify how well Kalman filter estimates (nowcasts) and Kalman smoother estimates (backcasts) track the truth. Pagan and Robinson (2022) and Buncic, Pagan and Robinson (2023) use simulated correlation coefficients between shocks and the Kalman smoother for assessing the quality of state space models. We use a more efficient quasi-analytical approach that allows us to unearth connections to estimation uncertainty and can be applied to levels and changes in stationary and non-stationary models.

Our measure of reliability – invertibility in the context of the Kalman Filter, recoverability in the context

of the Kalman Smoother – is an ex post measure: Given the observed data, how well can the econometrician estimate the current or past state? It is the flipside of predictability measures in the forecasting literature, e.g. Diebold and Kilian (2001). Diebold and Kilian (2001) contribute to the forecasting literature by extending predictability measures to non-stationary settings by focusing on the conditional variances. We borrow this insight from their work and apply it to reliability.

In settings with unit roots, we show under mild regularity conditions the Kalman filter estimate is cointegrated with the truth. This result is a special case of Engle and Yoo (1987) on forecasting. Since our reliability measure converges to unity at long horizons in unit-root settings, we suggest assessing the model dynamics at these horizons by evaluating the predictive accuracy of the state space model for the observables.

2 Environment and Kalman filtering recursions

Our environment is that of standard linear state-space models. Here, we use a formulation similar to Fernández-Villaverde et al. (2007). Namely, the (possibly hidden) state variables x_t load on their lagged value and a shock vector ϵ_t. Similarly, observables y_t load on the past state vector x_{t-1} and current shocks ϵ_t.

$$x_t = Ax_{t-1} + B_t\epsilon_t \tag{2.1a}$$

$$y_t = Cx_{t-1} + D_t\epsilon_t \tag{2.1b}$$

When $BD' \neq 0$, the innovations to x_t and y_t are uncorrelated. In general, however, they are correlated. The primitive shocks ϵ_t are iid Gaussian: $\mathcal{N}(0, I)$. A may be a stable matrix or may have one or more unit roots.

For stationary models, we demean y_t. For non-stationary models, we specify a prior $x_0 \sim \mathcal{N}(\hat{x}_0, \tilde{B}_0\tilde{B}_0')$, where $\tilde{B}_0$ is typically a diagonal with large entries. B_t, D_t are time varying to capture COVID.

2.1 Kalman filtering recursion

The Kalman smoother is optimal in this context, and the filtering equations are standard (e.g. Fernández-Villaverde et al. (2007)):

$$x_t|\mathscr{F}_{t-1}, y_t \sim N\left(\underbrace{A\hat{x}_{t-1|t-1} + K(y_t - C\hat{x}_{t-1|t-1})}_{\equiv\hat{x}_{t|t}}, \Omega_{t|t-1} - KV_{t-1}K_{1,t}'\right) \tag{2.2}$$

$$V_t = C\Omega_{t-1|t-1}C' + DD' \tag{2.3}$$

$$K = (A\Omega_{t-1|t-1}C' + BD')V_t^{-1}. \tag{2.4}$$

$$\Omega_{t|t-1} = A\Omega_{t-1|t-1}A' + BB' \tag{2.5}$$
$$\Omega_{t|t} = \Omega_{t|t-1} - KV_tK' \tag{2.6}$$

and hence:

$$\begin{aligned}\hat{x}_{t|t} &= A\hat{x}_{t-1|t-1} + K(y_t - C\hat{x}_{t-1|t-1}) \\ &= A\hat{x}_{t-1|t-1} + KC(x_{t-1} - \hat{x}_{t-1|t-1}) + KD\epsilon_t \\ &= (A - KC)\hat{x}_{t-1|t-1} + KCx_{t-1} + KD\epsilon_t. \end{aligned} \tag{2.7}$$

Stacking the laws of motion for the state and the filtered estimate:

$$\xi_t \equiv \begin{bmatrix} x_t \\ \hat{x}_{t|t} \end{bmatrix} = \underbrace{\begin{bmatrix} A & 0 \\ KC & A - KC \end{bmatrix}}_{\tilde{A}} \xi_{t-1} + \underbrace{\begin{bmatrix} B \\ KD \end{bmatrix}}_{\tilde{B}} \epsilon_t \tag{2.8}$$

In general, we can use (2.8) to calculate $H < \infty$ step ahead covariances and autocovariances for ξ_t, from which we can deduce autocorrelations and correlations conditional on the time t information $\mathscr{F}_t$.

To preview these calculation, consider the case of a stable matrix A. Then we can use the above formulation to calculate also the unconditional variance and unconditional correlation. Equation (2.8) implies the following unconditional variance-covariance matrix:

$$\mathrm{Var}[\xi_t] = \sum_{h=0}^{\infty} \tilde{A}^h \begin{bmatrix} BB' & B(KD)' \\ (KD)B' & (KD)(KD)' \end{bmatrix} (\tilde{A}')^h \tag{2.9}$$

$$\Leftrightarrow \quad \mathrm{vec}(\mathrm{Var}[\xi_t]) = (I - \tilde{A} \otimes \tilde{A})^{-1} \mathrm{vec} \begin{bmatrix} BB' & B(KD)' \\ (KD)B' & (KD)(KD)' \end{bmatrix} \tag{2.10}$$

The correlation of the filtered state estimate and the truth follows from (2.9) or, equivalently, (2.10).

From (2.8) we can calculate the law of motion of the h-period ahead estimation error z_{t+h}:

$$z_{t+h} = x_{t+h} - \hat{x}_{t+h|t+h} = (A - KC)z_{t+h-1} + (B - KD)\epsilon_{t+h}$$

To see how latent state and its filter-based estimate jointly behave in a simple setting, consider the following simple example:

$$\begin{aligned} x_t &= \rho x_{t-1} + [1,0]\epsilon_t, \\ y_t &= x_{t-1} + [0,\sigma]\epsilon_t. \end{aligned}$$

The setting above, clearly, covers both the stationary case ($|\rho| < 1$) and the unit root case ($\rho = 1$). The steady state filter for the model is such:

$$\begin{aligned} \Omega_{t|t} &= \frac{1-\sigma^2(1-\rho^2)}{2} + \sqrt{\left[\frac{1-\sigma^2(1-\rho^2)}{2}\right]^2 + \sigma^2} \\ K &= \frac{\rho\Omega_{t|t}}{\Omega_{t|t}+\sigma^2} \in (0,1) \\ \tilde{A} &= \begin{bmatrix} \rho & 0 \\ \frac{\rho\Omega_{t|t}}{\Omega_{t|t}+\sigma^2} & \frac{\rho\sigma^2}{\Omega_{t|t}+\sigma^2} \end{bmatrix} \end{aligned}$$

Note that the eigenvalues of $\hat{A}$ are ρ and $\frac{\rho\sigma^2}{\Omega_{t|t}+\sigma^2}$, the latter always less than one even when $\rho = 1$. When $\rho = 1$ x_t and its estimate $\hat{x}_{t|t}$ share a unit root and their difference is stationary:

$$\begin{aligned} \hat{x}_{t|t} &= (1-K)\hat{x}_{t-1} + Ky_t \\ &= (\rho - K)\hat{x}_{t-1} + Kx_{t-1} + K\sigma\epsilon_{2,t} \end{aligned}$$

Hence, when $\rho = 1$, x_t and $\hat{x}_{t|t}$ are cointegrated, with $1 - K = \frac{\sigma^2}{\Omega_{t|t}+\sigma^2}$ the speed of adjustment.

Engle and Yoo (1987) have shown that in non-stationary settings, the forecast of a non-stationary variable is co-integrated with the truth. This also applies in our setting under mild regularity conditions, as stated in the following theorem:

Theorem 1. *If A has at least one unit root, if (A, B) is completely stabilizable and if (A, C) is completely detectable, then $\hat{x}_{t|t}$ and x_t are cointegrated with a stable "speed of adjustment" matrix $A - KC$.*

The proof of the theorem as well as definitions of stabilizability and detectability are given in Appendix A. Intuitively, a system that is completely stabilizable makes sure that the non-stationary states are subject to shocks. Complete detectability means that the non-stationary components of the system are observable through measurements. Without these conditions the filter variance could grow without bound or the filter could stop learning about some states.

We now show that the above reliability measures are – with the exception of the unconditional correlation – also applicable in systems with unit roots. In a nutshell, we show that under mild regularity conditions the state estimates are cointegrated with the truth and the estimation error is thus station-

ary. Finite-horizon (conditional) correlations and the half-life are thus well defined. The unconditional correlation, however, mechanically converges to unity and only the speed of convergence is of interest.

2.2 Extensions

Kalman Smoother. The Kalman Filter estimate is the best estimate using current and past data. The Kalman smoother uses future data as well and thus speaks to the recoverability. While it is typically used to smooth the entire history of state estimates, we may want to know how much the state estimates improve with the benefit of several periods of hindsight. We are thus not interested in the entire sequence, but rather in what (Harvey, 1990, p. 150) calls "fixed-lag smoothing".

For illustrative purposes, we derive the one-lag smoother below and show that the formal treatment is analogous to that using the Kalman filter:

$$\begin{aligned}\hat{x}_{T-1|T} &= \hat{x}_{T-1|T-1} + J_1(\hat{x}_{T|T} - A\hat{x}_{T-1|T-1}) \\ &= (I - J_1 A)\hat{x}_{T-1|T-1} + J_1\hat{x}_{T|T}. \qquad (2.11)\\ J_1 &= \Omega_{t|t}A'(A\Omega_{t|t}A' + BB')^{-1} \\ &= \Omega_{t-1|t-1}A'\Omega_{t|t-1}^{-1} \qquad (2.12)\end{aligned}$$

Stacking these equations with the Kalman filter system (2.8), we obtain an analogous augmented system:

$$\xi_t^+ \equiv \begin{bmatrix} x_t \\ \hat{x}_{t|t} \\ x_{t-1} \\ \hat{x}_{t-1|t} \end{bmatrix} = \begin{bmatrix} A & 0 & 0 & 0 \\ KC & A-KC & 0 & 0 \\ I & 0 & 0 & 0 \\ J_1KC & I-J_1KC & 0 & 0 \end{bmatrix} \xi_{t-1}^+ + \begin{bmatrix} B \\ KD \\ 0 \\ J_1KD \end{bmatrix} \epsilon_t \qquad (2.13)$$

Using this definition, the definition of the smoother error follows as:

$$z_{t-1}^{sm} \equiv x_{t-1} - \hat{x}_{t-1|t} = \begin{bmatrix} 0 & 0 & I & -I \end{bmatrix} \xi_{t-1}^+ = (I - J_1KC)(x_{t-1} - \hat{x}_{t-1|t-1}) + J_1KD\epsilon_t$$

Forward this by one period:

$$z_t^{sm} = (I - J_1KC)(x_t - \hat{x}_{t|t}) + J_1KD\epsilon_{t+1} = (I - J_1KC)z_t + J_1KD\epsilon_{t+1}$$

From these expressions, we can follow the same steps as above to derive reliability measures.

Equation (2.13) shows that evaluating the Kalman-smoothed estimates analytically requires augmenting

the state space. While these derivations are useful for illustrative purposes, in practice it is easier to augment the original state vector x_t with lagged copies of the states to implement the fixed-lag Kalman smoothing.

2.3 Comparing Posterior Variances: Level vs. Difference

Inferring the levels of latent variables is often difficult when the latent variables are non-stationary or highly persistent. In many cases, the observable data tend to contain more information about short-run changes than about the long-run position of the underlying state. When the signals primarily reflect changes rather than levels, the estimation of the level is difficult, and its posterior uncertainty can remain large even after many observations. By contrast, inference about differences is often much sharper, since these components are directly linked to the innovations driving the system. Even when the signal conveys information about both the level and the change, the level is harder to pin down because past shocks accumulate over time and are only imperfectly corrected by new data. As a result, estimates of changes can be more reliable than estimates of levels.

We now present a simple model to illustrate why changes in a latent variable can often be inferred more precisely than its level. Suppose that the latent process x_t follows a univariate random walk, and that $\epsilon_t = [\epsilon_{1,t}, \epsilon_{2,t}]'$ is a vector of idiosyncratic shocks:

$$\begin{aligned} \Delta x_t &= [\sigma_x, 0]\epsilon_t && (2.14) \\ y_t &= cx_{t-1} + [d, \sigma_y]\epsilon_t \\ &= \gamma x_t + (1-\gamma)\Delta x_t + \sigma_y \epsilon_{2,t}. && (2.15) \end{aligned}$$

where $c = \gamma$ and $d = \sigma_x$. Here, y_t is the observable signal that contains information about the latent state x_t. The parameter γ controls how strongly the signal loads on the level of the state as opposed to its short-run change. When γ is close to one, the signal primarily reflects the level; when γ is close to zero, it mostly reflects the change.

Consider first the special case with $\gamma = 0$, so that $y_t = \Delta x_t + \text{noise}_t$. In this setting, the signal contains information only about short-run innovations in the state, not about its level. Intuitively, observing y_t helps us learn about the most recent change in x_t, but provides no information about where x_t actually started. As a result, the initial level remains completely unidentified, and its posterior uncertainty grows without bound.

More generally, when the signal depends both on the level and the change in the latent state, we can use the Kalman filter to update our belief about x_t. The posterior variance of the state, denoted Ω'_x, measures how much uncertainty remains about the level of x_t after seeing y_t. We can compute the posterior variance Ω'_x of x_t as

$$\Omega'_x = \frac{\Omega_x[\sigma_x^2(1-\gamma)^2 + \sigma_y^2] + \sigma_x^2\sigma_y^2}{\gamma^2\Omega_x + \sigma_x^2 + \sigma_y^2} = \begin{cases} \frac{\Omega_x[\sigma_x^2+\sigma_y^2]+\sigma_x^2\sigma_y^2}{\Omega_x+\sigma_x^2+\sigma_y^2} & \gamma = 1 \\ \nearrow \infty & \gamma \searrow 0. \end{cases}$$

Analytically, Ω'_x depends on the parameters σ_x, σ_y, γ, and the prior uncertainty Ω_x. When $\gamma = 1$, the signal contains direct information about the level, and the Kalman filter can gradually learn x_t over time. When $\gamma \to 0$, however, the signal conveys little or no information about the level, and the posterior variance of x_t diverges, meaning that the level cannot be reliably inferred.

Note, however, that when $\gamma = 0$, inference for $\Delta x_t = \sigma_x \epsilon_t$ is straightforward with posterior variance $\Omega_{\Delta x} = \frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}$. This is smaller than the level uncertainty Ω_x even if $\gamma = 1$.[3]

More generally, we can consider inference on h-step differences $x_t - x_{t-h}$. Depending on the signal structure and the persistence of the state, we can expect these changes to be inferred more precisely than the level. To consider such long difference, we again augment the state space. We can thus consider the reliability of level Kalman-filter estimates, fixed-lag Kalman-smoother estimates, or long differences in a unified framework.

3 Reliability measures

How closely does a state estimate track the true latent variable? By default, the literature uses the posterior variance to provide a quantitative answer to this question. We complement this perspective by introducing a dynamic conditional forecast-error-variance decomposition and a related half-life statistic. These allow us to better understand what shapes the width of the probability intervals surrounding the state estimates. In addition, we provide a novel perspective: How closely a filtered estimate tracks the true latent variable can be measured by an R^2 statistic. This measures how well the econometrician can "invert" the data to back out the current state. We develop this R^2 statistic next before connecting it to the well-known posterior uncertainty and its drivers.

If we could regress a true (stationary, scalar) state $x_{i,t}$ on its filtered estimate, a standard goodness-of-fit measure is the R^2 of the following univariate regression:

$$x_{i,t} = \alpha + \beta \hat{x}_{i,t|t} + \nu_{i,t}. \tag{3.1}$$

Since $\hat{x}_{t|t}$ is the best linear estimator given information $\boldsymbol{y}^t$ and it is unbiased – so that $\beta = 1$ and $\alpha = 0$. In addition, the variance of the regression error $\nu_{i,t}$ is the $i-^{th}$ diagonal element of the Kalman filter estimate variance $\Omega_{t|t}$. The R^2 is then one minus the ratio of the residual sum of squares (RSS) to the total sum

[3] To see this, set $\gamma = 1$. Then the steady state Ω_x satisfies $\Omega_x = \frac{\sigma_y^2(\Omega_x+\sigma_x^2)}{\Omega_x+\sigma_x^2+\sigma_y^2}$. Cross-multiplying gives $\Omega_x(\Omega_x + \sigma_x^2) = \sigma_x^2\sigma_y^2$. This equation is strictly increasing in Ω_x. We now show that $\Omega_x \leqslant \Omega_{\Delta x}$ would imply that $\Omega_x(\Omega_x+\sigma_x^2) < \sigma_x^2\sigma_y^2$, a contradiction. Setting $\Omega_x = \Omega_{\Delta x}$ in $\Omega_x(\Omega_x + \sigma_x^2)$ yields: $\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2} \cdot \left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2} + \sigma_x^2\right) = \frac{\sigma_x^4\sigma_y^2(\sigma_x^2+2\sigma_y^2)}{(\sigma_x^2+\sigma_y^2)^2} < \sigma_x^2\sigma_y^2$ We conclude that $\Omega_x > \Omega_{\Delta_x}$

of squares (TSS). The TSS divided by the sample size converges to variance, and similarly for the RSS. Thus, the sum of squares are simply T times the posterior and prior variance of $x_{i,t}$:

$$\begin{aligned} R^2 &= 1 - \frac{\text{RSS}/T}{\text{TSS}/T} = 1 - \frac{\Omega_{ii}}{\Sigma_{ii}} \\ &= \frac{[A(\Sigma - \Omega)A' + KVK']_{ii}}{[A\Sigma A' + BB']_{ii}} = \frac{\text{ESS}/T}{\text{TSS}/T}. \end{aligned} \tag{3.2}$$

where Σ denotes the unconditional state variance (the solution of $\Sigma = A\Sigma A' + BB'$) we assume that the state estimate uses the steady-state Kalman Filter. Alternatively, the equality follows as the $T \to \infty$ limit. The second line re-writes the identity using the explained sum-of-squares. The Kalman filter explains the variation in x_t partly by reducing the prior variance from Σ to Ω and by learning from the latest signal with variance V.

The interpretation of the R^2 is straightforward: On average, which fraction of the movements in the true state is explained by the filtered state?

This R^2 measure is intimately connected to the probability bands surrounding the state estimate. Given a set of parameters, the posterior uncertainty is Ω_{ii}. Clearly, as $\Omega_{ii} \searrow 0$, $R^2 \nearrow 1$. And if the signals are near-useless, then $\Omega_{ii} \nearrow \Sigma_{ii}$ and the R^2 approaches zero.[4]

There is, however, a limitation to the unconditional R^2-measure (3.2): It only applies to stationary systems. In the presence of unit roots, the unconditional regression mechanically produces an R^2 equal to one, since $R^2 = 1 - \frac{\Omega_{ii}}{\Sigma_{ii}}$, and Ω_{ii} diverges.

3.1 Conditional and unconditional correlation and R^2 analogues

The (infeasible) regression (3.1) regresses the state on the filtered estimate – it reflects the unconditional predictive performance. The state-space system also allows us, however, to evaluate a conditional counterpart: Which fraction of the forecast error variance in $x_{i,t+h}$ given $\boldsymbol{y}^t$ can we explain using the information flow between times t and $t+h$? This is the R^2 in regression (3.1) where each side is expressed in terms of deviations from their projection on the information set including observables up to h-periods before:

$$\begin{aligned} &x^{res}_{i,t+h} = x^{res}_{i,t+h|t+h} + \nu_{i,h,t}, \\ &x^{res}_{i,t+h} \equiv x_{i,t+h} - \mathbb{E}[x_{i,t+h}|\boldsymbol{y}^t] \qquad \hat{x}^{res}_{i,t+h|t+h} = \hat{x}_{i,t+h|t+h} - \underbrace{\mathbb{E}[\hat{x}_{i,t+h|t+h}|\boldsymbol{y}^t]}_{=\hat{x}_{i,t+h|t}}. \end{aligned} \tag{3.3}$$

Compared to the unconditional R^2, the conditional R^2 in this regression has a more subtle interpretation. The variance that we are trying to explain is the unanticipated change in the true state from time t to $t+h$.

[4] In our model, R^2 and Ω are functions of the same set of parameters. If the persistence increases or the variance of the state innovation increases, it is possible that both Ω_{ii} and R^2 increase. In contrast, when the signal noise rises, we would typically expect Ω_{ii} to rise and R^2 to fall.

Our predictor is the sufficient statistic that summarizes the information received from the observables over the same period. For example, if $\hat{x}_{t|t}$ is a nowcast, it asks how much of the unanticipated change in x_{t+h} is explained by the new information in the nowcast, which is $\hat{x}^{res}_{t+h|t+h} = \hat{x}_{i,t+h|t+h} - \hat{x}_{i,t+h|t}$.

While we could compute the R^2 in the unconditional regression using the unconditional variance given in (2.10), the conditional regression requires conditioning on the observed data. Since observing $\boldsymbol{y}^t$ leaves us with $\text{Var}[x_t|\boldsymbol{y}^t] = \Omega_{t|t}$, this information set gives rise to the following conditional variance-covariance matrices:

$$\text{Var}[\xi_{t+h}|y^t] = \tilde{A}^h \begin{bmatrix} \Omega_{t|t} & 0 \\ 0 & 0 \end{bmatrix} (\tilde{A}')^h + \sum_{j=1}^{h} \tilde{A}^{j-1} \tilde{B}\tilde{B}' (\tilde{A}')^{j-1}, \qquad h \geqslant 0. \tag{3.4}$$

If $\tilde{A}$ is stable, (3.4) converges to an unconditional covariance, from which we can deduce the conditional correlation. The conditional correlation tells us how closely our filtered estimates are associated with the truth h quarters into the future.

For x_{t+h} itself, a simpler, more explicit expression is available. Since $\tilde{A}$ is block-lower-triangular, a simpler expression is available for the conditional variance if x_{t+h}:

$$\text{Var}[x_{t+h}|y^t] = A^h \Omega_{t|t} (A')^h + \sum_{j=1}^{h} A^{j-1} BB' (A')^{j-1} \tag{3.5}$$

These variance measures are useful because they allow us to compute all objects implied by the population regression functions, in particular regression coefficients and R^2 measures, as the following Lemma illustrates.

Lemma 3.1 (Regression properties.)**.** *If $h \geqslant 1$, (a) the coefficient β_h in the population regression (3.3) equals 1 and (b) the squared conditional correlation between the true state x_{t+h} and the state estimate $\hat{x}_{t+h|t+h}$ equals the conditional R^2. Here, the conditioning set may include the set of all observables.*

Proof. For $h \geqslant 1$:

$$\begin{aligned}
\text{Cov}[x_{t+h+1}, \hat{x}_{t+h+1|t+h+1}|y^t] &= \mathbb{E}[(x_{t+h+1} - \hat{x}_{t+h+1|t+h+1} + \hat{x}_{t+h+1|t+h+1} - \mathbb{E}[\hat{x}_{t+h+1|t+h+1}|y^t]) \\
&\qquad \times (\hat{x}_{t+h+1|t+h+1} - \mathbb{E}[\hat{x}_{t+h+1|t+h+1}|y^t])'|y^t] \\
&= \mathbb{E}[\mathbb{E}[(x_{t+h+1} - \hat{x}_{t+h+1|t+h+1})(\hat{x}_{t+h+1|t+h+1} - \mathbb{E}[\hat{x}_{t+h+1|t+h+1}|y^t])'|y^{t+h+1}]|y^t] \\
&\qquad + \text{Var}[\hat{x}_{t+h+1|t+h+1}|y^t] \\
&= \text{Var}[\hat{x}_{t+h+1|t+h+1}|y^t]
\end{aligned}$$

using the law of iterated expectations (LIE). Since $\beta_h = \frac{\text{Cov}[x_{t+h+1}, \hat{x}_{t+h+1|t+h+1}|y^t]}{\text{Var}[\hat{x}_{t+h+1|t+h+1}|y^t]} = 1$, part (a) follows.

But then using that Cov = Var[filter], we have that

$$\text{corr} = \frac{\text{Cov}[\text{Truth}, \text{Filter}]}{\sqrt{\text{Var}[\text{Truth}]\,\text{Var}[\text{Filter}]}} = \frac{\sqrt{\text{Var}[\text{Filter}]}}{\sqrt{\text{Var}[\text{Truth}]}} = \text{rel sd} \tag{3.6}$$

Thus, the squared correlation equals the share of the variance explained by the filter. □

To derive the conditional R_h^2, note that $\hat{x}_{t+h|t+h}$ is a sufficient statistic for x_t, summarizing all information in y^t. Thus, conditional on $\hat{x}_{t+h|t+h}$, the residual variance in regression (3.3) is always Ω when using the steady state Kalman filter. Thus:

$$R_h^2 = 1 - \frac{\text{RSS}/T}{\text{TSS}/T} = 1 - \frac{\Omega_{ii}}{\text{Var}[x_{t+h}|y^t]_{ii}} \tag{3.7}$$

For $h = 1$, $\text{Var}[x_{t+h}|y^t]_{ii} = A\Omega A' + BB'$ and we can write:

$$\begin{aligned} R_1^2 &= 1 - \frac{\Omega_{ii}}{[A\Omega A' + BB']_{ii}} = \frac{[A\Omega A' + BB']]_{ii} - [A\Omega A' + BB' - KVK']_{ii}}{[A\Omega A' + BB']_{ii}} \\ &= \frac{[KVK']_{ii}}{[A\Omega A' + BB']_{ii}} = \frac{\text{ESS}/T}{\text{TSS}/T}. \end{aligned} \tag{3.8}$$

In this conditional regression, the explained sum of squares is thus simply equal to the quadratic form of the Kalman gain K and the variance of the innovation to the observables. As before, when the state is recovered near perfectly, the R_h^2 approaches 1 for any h. As Ω_{ii} grows, the R_h^2 approaches zero.

Note that this measure allows us to quantify the value of knowing the truth in terms of an increment in R_h^2:

$$x_{i,t+h}^{res} = \tilde{\alpha}_h + \tilde{\beta}_h \hat{x}_{i,t+h|t+h}^{res} + \gamma_h x_{i,t}^{res} + \tilde{\nu}_{i,h,t}, \tag{3.9}$$

where $\hat{x}_{i,t}^{res} = \hat{x}_{i,t} - \mathbb{E}[\hat{x}_{i,t}|\boldsymbol{y}^t]$.[5] This equals the state estimation error z_t defined below. The R^2 in this multivariate regression is now:

$$R_{h,x(i,t)}^2 = \text{corr}(x_{i,t+h}, \mathbb{E}[x_{i,t+h}|x_{i,t+h|t+h}, x_{i,t}]|y^t)^2 = \frac{\Sigma_{x_{i,t+h}W}\Sigma_{WW}^{-1}\Sigma_{Wx_{t+h}}}{\Sigma_{x_{t+h},x_{t+h}}} \tag{3.10}$$

where $W = [\hat{x}_{i,t+h|t+h}, x_{i,t}]'$. For calculating this measure, we also need the covariance between $\xi_t = [x_t', \hat{x}_{t|t}']'$ and ξ_{t+h} conditional on y^t. Since the innovations are independent, the covariance is entirely due to state uncertainty:

$$\text{Cov}[\xi_t, \xi_{t+h}] = \begin{bmatrix} \Omega_{t|t} & 0 \\ 0 & 0 \end{bmatrix} (\tilde{A}')^h.$$

[5] Here, $\tilde{\beta}_h \neq 1$ in general.

By comparing the different R^2 measures, we can define the value of knowing the initial state with certainty as the difference between the R^2 measures in (3.10) and (3.7):

$$\text{Increase in predictability of } x_{i,t+h} \text{ due to } x_{i,t} = R^2_{h,x(i,t)} - R^2_h.$$

It is straightforward to generalizing this calculation to having a one-time noisy observation of $x_{i,t}$ from outside the model. In practice, this could reflect, for example, subjective knowledge or estimates from micro-data.

One immediate implication is that the conditional $\lim_{h\to\infty} R^2_h$ converges to unity for non-stationary variables under mild regularity conditions.

Theorem 3.2. *Suppose $x^{ur}_t \subset x_t$ has a unit root and each of its components is non-degenerate. Assume that (A, B) is completely stabilizable and (A, C) is completely detectable, then the R^2 in a regression of any element of x^{ur}_{t+h} on its Kalman filter estimate $\hat{x}^{ur}_{t+h|t+h}$ and the initial state vector estimate $\hat{x}_{t|t}$ approaches 1 as $h \to \infty$ when initialized with the stationary Kalman filter.*

Proof. Under the conditions of the theorem, the speed of adjustment matrix $A - KC$ is stable due to Theorem 1. Hence, the conditional variance $\Omega = \text{Var}[x_t|y^t]$ is finite. Note that $\text{Var}[x_{t+h}|y^t] = A^h\Omega(A^h)' + \sum_{\ell=1}^{h} A^{\ell-1}BB'(A^{\ell-1})'$. Because each component of x^{ur}_t is non-degenerate and has a unit-root, the diagonal of $\text{Var}[x^{ur}_{t+h}|y^t]$ increases without bound. However, after accounting for y^h, its variance is $\Omega_{ur,ur}$, a finite PSD matrix. The R^2 for some variable $x_{i,t} \subset x^{ur}_t$ at horizon h is given by $R^2_h \equiv 1 - \frac{\Omega_{i,i}}{\text{Var}[x_{i,t+h}|y^t]}$. Since all elements of Ω are finite, but all elements of $\text{Var}[x^{ur}_{t+h}|y^t]$, including $\text{Var}[x_{i,t+h}|y^t]$, diverge, we have that $R^2_h \to 1$. □

Since the loading of the true state on the state estimate approaches +1 and is thus positive, the following corollary is immediate:

Corollary 3.3. *Under the conditions of Proposition 3.2, the correlation of $\hat{x}_{t+h|t+h}$ with x_{t+h} conditional on y^t approaches +1.*

Intuitively, under our regularity conditions the Kalman Filter reduces the near-infinity conditional variance of x_{t+h} to the finite Ω, the residual sum of squares becomes vanishingly small.

Note that this result contrasts with the limit for stationary variables. For stationary variables, $\lim_{h\to\infty} R^2_h = R^2$, which is generally smaller than one.

The contrast between stationary and non-stationary models highlights the importance of assessing whether the model dynamics are correctly specified. While short-run dynamics may be well described by a (near-)unit root, perhaps there is mean-reversion over longer horizons? Our result suggests that longer-horizon reliability measures can be sensitive to potential misspecification. We therefore suggest in Section 6 a careful assessment of the observable implications of the state-space model to discipline the model dynamics.

We could also compare the predictability of state $x_{i,t+h}$ under different information sets. For stationary models using the unconditional R^2 this is straightforward because the explanatory power is measured relative to a common denominator. For the conditional R^2_h, however, the conditional variance that we are trying to explain also varies across models. Comparing conditional R^2_h measures under different information sets can thus be misleading. With nested information sets, the differences in the Ω variance matrices are positive semi-definite. The denominator of R^2_h is, however, a quadratic form that needn't satisfy the same ranking as that of the variances.[6] When using the conditional R^2_h for comparing information sets we therefore recommend using the larger of the conditional variances as the denominator.

3.2 Dissecting state uncertainty: Half-life, and variance decomposition

The state uncertainty Ω is a key object in the steady state Kalman filter, governing both the predictability as measured by our R^2 measures but also measuring the uncertainty about the latent state. We argue that it is useful to look at two related objects that shape Ω: The half-life of state estimation errors and the contribution of uncertainty about the initial state.

Iterating forward h periods gives

$$z_{t+h} = (A - KC)^h (x_t - \hat{x}_{t|t}) + \sum_{j=1}^{h} (A - KC)^{j-1} (B - KD)\epsilon_{t+h-j+1} \tag{3.11}$$

The first term on the right-hand side of (3.11) reflects the decay of the initial estimation error, while the second accumulates the effect of new shocks.

At steady state, the variance of the estimation error is constant and equal to the posterior variance, that is $\text{Var}[z_{t+1}|\mathscr{F}_t] = \Omega_{t|t} = \Omega$. Intuitively, since we initialize the filter in steady state and $z_{t+1} = x_{t+1} - \mathbb{E}[x_{t+1}|\mathscr{F}_t + 1]$, the increase in uncertainty from future shocks is exactly offset by the reduction in uncertainty due to future signals.

Lemma 2. *With the steady state Kalman filter, the variance of the state estimation error z_t equals Ω, the posterior variance of the state in t given time t information.*

Proof. Let V be the ex ante variance of the measurement. Focus on the time-less Kalman filter. From the definition of the Kalman gain: $KV = A\Omega_{t|t}C' + BD'$ so that $KVK' = A\Omega_{t|t}(KC)' + B(KD')$. Let $A_{KC} = A - KC$ and $B_{KD} = B - KD$. Then $\text{Var}[z_{t+1}|\mathscr{F}_t] = A_{KC}\Omega_{t|t}A'_{KC} + \tilde{B}_{KD}B'_{KD}$. Multiplying this equation out we get:

$$\begin{aligned}\text{Var}[z_{t+1}|\mathscr{F}_t] &= A\Omega_{t|t}A' - A\Omega_{t|t}(KC)' - KC\Omega_{t|t}A' + KC\Omega_{t|t}(KC)' + BB' - B(KD)' + KDB' - KDD'K' \\ &= A\Omega_{t|t}A' - 2KVK' + B(KD)' + KDB' + KC\Omega_{t|t}(KC)' + BB' - B(KD)' + KDB' - KDD'K'\end{aligned}$$

[6] There are, however, important special cases. For example, if x_t is a scalar, then for nested information sets differences in Ω translate to differences in R^2_h. The same is true for $x_{i,t} = e_i'x_t$ if e_i' is a left eigenvector of A.

$$= A\Omega_{t|t}A' + BB' - 2KVK' + K\underbrace{(C\Omega_{t|t}C' + DD')}_{\equiv V}K' = A\Omega_{t|t}A' + BB' - KVK'.$$

The last expression equals $\Omega_{t+1|t+1}$ and Ω in steady state. □

Thus the unconditional variance of the VAR(1) in (3.11) is the state uncertainty. Computationally, we can obtain the unconditional variance as the limit of the conditional variance $\text{Var}[z_{t+h}|y^t]$ which we can express as:

$$\text{Var}[z_{t+h}|y^t] = (A-KC)^h\Omega((A-KC)')h + \sum_{\ell=1}^{h}(A-KC)^{\ell-1}(B-KD)(B-KD)'((A-KC)')^{\ell-1}$$

The first term captures the contribution of *initial uncertainty* (the uncertainty inherited from time t), while the second term captures the accumulation of new noise shocks between t and $t+h$.

Since we also have that $\text{Var}[z_{t+h}|y^t] = \Omega$, we can express the fraction of current uncertainty that stems from the initial state as:

$$FEVD_{i,h} = 100 - 100 \times \frac{\left[\sum_{\ell=1}^{h}(A-KC)^{\ell-1}(B-KD)(B-KD)'((A-KC)')^{\ell-1}\right]_{ii}}{[\Omega_{t|t}]_{ii}}. \tag{3.12}$$

This "forecast error variance decomposition" (FEVD) quantifies how much of the posterior uncertainty is due to the persistence of prior uncertainty rather than new shocks.

What shapes the contribution of the initial state uncertainty and how much the effect of shocks accumulates is, of course, how quickly $A-KC$ decays. In the scalar case, this is transparent, as $\Omega = (B-KD)(B-KD)'/(1-(A-KC)^2)$: A higher persistence parameter $A-KC$ implies a larger steady-state variance Ω. In general:

$$\text{vec}(\Omega) = (I - (A-KC)\otimes(A-KC))^{-1}\text{vec}((B-KD)(B-KD)'). \tag{3.13}$$

This motivates analyzing the autocorrelation function for state s from the autocovariance induced by (3.11), which can be shown to be:

$$\text{Cov}[z_{t+h}, z_{t+1}|\mathscr{F}_t] = (A-KC)^{h-1}\text{Cov}[z_{t+1}, z_{t+1}|\mathscr{F}_t], \tag{3.14}$$

where

$$\text{Var}[z_{t+1}|\mathscr{F}_t] = \text{Cov}[z_{t+1}, z_{t+1}|\mathscr{F}_t] = (B-KD)(B-KD)' + (A-KC)\Omega_{t|t}(A-KC).$$

The autocorrelation function (ACF) of the estimation error is thus:

$$ACF_{s,h} = \frac{e_s\big((A-KC)^h\Omega_{t+1|t}\big)e_s'}{e_s\Omega_{t+1|t}e_s'} \tag{3.15}$$

Definition 3. *The overall half-life of state j is $\tau_j^{1/2} \equiv \min h \geqslant 1 : ACF_{j,h} < \frac{1}{2}$ if such h exists, and $\tau_j^{1/2} = \infty$ otherwise.*

The half-life of state j is defined as the smallest horizon h at which the ACF drops below one-half. For the entire state vector, the largest eigenvalue of $A - KC$ determines the speed of convergence. For specific states, however, convergence can be faster.

As equation (3.13) makes clear, however, half-life and FEVD are not independent measures of state uncertainty but rather speak to the origins of state uncertainty: The half-life measures **how long** estimation errors survive; the FEVD measures **how much** of today's uncertainty is inherited from the past. Note that these concepts can be applied equally to stationary and non-stationary models.

3.3 Relationship between uncertainty, predictability, and half-life

Posterior uncertainty Ω, the unconditional R^2, and the half-life $\tau^{1/2}$ are jointly determined. Here, we focus on the scalar case in the stationary setting. We furthermore assume that B and D are scalars.

The three measures are tied together by the scalar steady-state Riccati equation $\Omega = \frac{(B-KD)^2}{1-(A-KC)^2}$, which can be written as:

$$\underbrace{R^2 = 1 - \frac{\Omega}{\Sigma}}_{\text{unconditional predictability}} \qquad \underbrace{\Omega = \frac{(B-KD)^2}{1-(A-KC)^2}}_{\text{persistence amplifies noise}} \qquad \underbrace{\tau^{1/2} = \frac{-\ln 2}{\ln|A-KC|}}_{\text{ACF decay rate}} \tag{3.16}$$

where $\Sigma = \frac{BB'}{1-A^2}$ is the unconditional variance.

The three measures are mutually constrained through the steady-state Riccati equation, which determines Ω and K jointly as functions of the structural primitives (A, B, C, D). Once solved, the Kalman gain K pins down $A - KC$ and $B - KD$, and the Riccati identity holds as a fixed-point condition. The joint constraint linking all three reliability measures is:

$$1 - R^2 = \frac{(1-A^2)}{B^2}\Omega = \frac{1-A^2}{(1-(A-KC)^2)}\frac{(B-KD)^2}{B^2}, \tag{3.17}$$

where $B - KD$ and $A - KC$ depend on the endogenous $K(\Omega)$. $A - KC$ can be written in terms of the half-life as $A - KC = 2^{-1/\tau^{1/2}}$. A change in any single structural parameter generally moves K, Ω, and Σ simultaneously, making the pairwise relationships between R^2, Ω, and $\tau^{1/2}$ parameter-dependent. For example, this implies that holding $(B-KD)^2$ fixed, a longer half-life always implies larger Ω. Also, holding Σ fixed, larger Ω always implies lower R^2. More generally, the three measures must be computed jointly from the structural parameters via the Riccati equation.

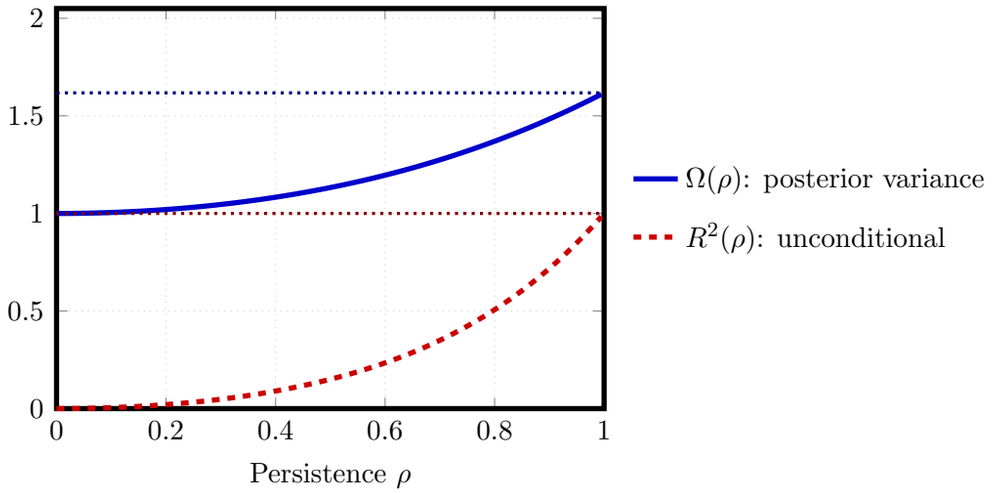


Figure 2: Numerical example showing how R^2 can rise along with the state variance: $x_t = \rho x_{t-1} + e_{1,t}, y_t = x_{t-1} + e_{2,t}$.

4 Frequentist & Bayesian Distributions of Reliability Measures

Thus far, we have computed reliability measures for a given parameter vector. When these measures are smooth functions of the underlying parameter vector θ, we can quantify the uncertainty around these reliability measures using parameter estimates. For example, if a frequentist estimator satisfies $\hat{\theta} - \theta \sim N(0, \text{AVar})$ and a reliability measure is expressed as $r(\theta)$, we can use the Delta method to derive its asymptotic distribution:

$$r(\hat{\theta}) - r(\theta) \sim N(0, \nabla(r)\text{AVar}\nabla(r)') \tag{4.1}$$

This distribution can be used to compute confidence intervals or to test hypotheses about reliability measures.

Similarly, from a Bayesian perspective, we can compute distributions of reliability measures. Let $f(\mathbf{y}^T|\theta)$ denote the likelihood function and $\pi(\theta)$ the prior over θ. The distribution of reliability measures $r(\theta)$ can be derived from either the prior distribution $\pi(\theta)$ or the posterior distribution $\pi(\theta|\mathbf{y}^T) \propto f(\mathbf{y}^T|\theta)\pi(\theta)$. Alternatively, if researchers have prior beliefs about the admissible range of a reliability measure, these can be incorporated by restricting the prior support accordingly. The evidence for or against such restrictions can be assessed using the Bayes factor:

$$\begin{aligned}\text{Bayes factor}_{\text{restricted}} &= \frac{\int f(\mathbf{y}^T|\theta) \frac{\pi(\theta)\mathbf{1}\{\theta \text{ satisfies restriction}\}}{\int \pi(\tilde{\theta})\mathbf{1}\{\tilde{\theta} \text{ satisfies restriction}\}d\tilde{\theta}} d\theta}{\int f(\mathbf{y}^T|\theta)\pi(\theta)d\theta} \\ &= \frac{\int \frac{f(\mathbf{y}^T|\theta)\pi(\theta)}{\int f(\mathbf{y}^T|\tilde{\theta})\pi(\tilde{\theta})d\tilde{\theta}} \mathbf{1}\{\theta \text{ satisfies restriction}\}d\theta}{\int \pi(\tilde{\theta})\mathbf{1}\{\tilde{\theta} \text{ satisfies restriction}\}d\tilde{\theta}}\end{aligned}$$

$$= \frac{\text{Posterior probability of restriction}}{\text{Prior probability of restriction}} \tag{4.2}$$

This is a special case of the Savage-Dickey ratio.

As one example, we now show that the conditional R^2 measure that we propose maps one-to-one into the smoothness parameter of the Hodrick and Prescott (1997, HP hereafter) filter. This parameter is routinely calibrated in the literature, suggesting that some researchers have prior beliefs over this parameter. Similarly, in the local level model R^2 can be tied to the smoothness of the trend and the literature has imposed priors over this smoothness.

4.1 Eliciting R^2 priors: HP filter example

The HP filter is a trend-cycle decomposition of the observable $y_t = \tau_t + c_t$, where τ_t is the trend and c_t is sometimes referred to as the cycle.[7] This model maps into our notation defining the state vector $x(t) = [\tau_t, \tau_{t-1}]'$ and the innovation vector $e(t) = [e_{\tau,t}.e_{c,t}]'$ with matrices:

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \sigma_\tau & 0 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & \lambda \end{bmatrix} \sigma_\tau$$

Now consider the one-step ahead conditional R^2 for the trend τ_t:

$$R_1^2 = 1 - \frac{\Omega_{11}}{[A\Omega A' + BB']_{11}} \tag{4.3}$$

Note that Ω is homogeneous of degree one in σ_τ^2, so that σ_τ^2 cancels. $\lambda = \frac{\sigma_c^2}{\sigma_\tau^2}$ governs the variance of the cyclical component to the trend component and equals the smoothing parameter.[8]

A more volatile trend corresponds to a higher R^2 in our framework. To show this, figure 3 plots equation (4.3) a range of smoothing parameters considered in the literature for annual to quarterly data. R_1^2 is strictly decreasing in λ: When λ is large, we either have a very smooth trend or, equivalently, relatively large measurement error. The state is thus hard to estimate. When λ is small, the HP-filter has the data track the trend closely and R_1^2 is high.

[7] Hamilton (2018) and others point out that this interpretation is problematic. We use the model here for illustrative purposes.

[8] See, for example, Proposition 1 in Hamilton (2018).

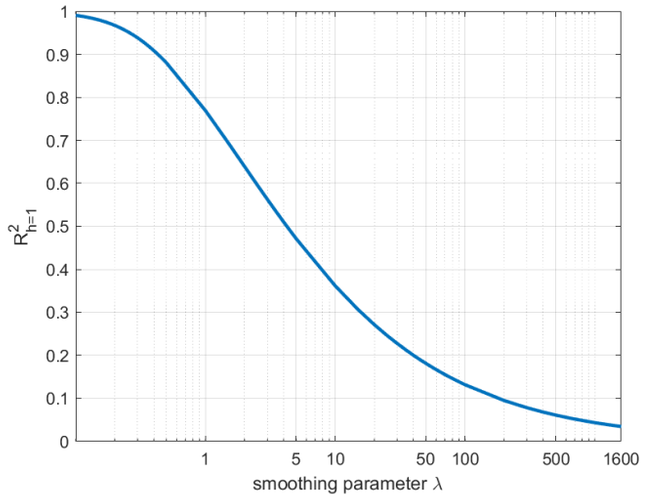


Figure 3: One-period ahead conditional R^2 for the trend estimate in the HP-filter case

In the HP-filter, our conditional R^2-measure thus maps one-to-one into the commonly calibrated smoothness parameter.

4.2 Eliciting R^2 priors: local level model

The same logic of the HP-example applies more broadly. For simplicity, consider the local level model: $x_t = x_{t-1} + \sigma_x \epsilon_{x,t}$ and $y_t = cx_t + \sigma_y \epsilon_{y,t}$. It is easy to show that when we introduce $\lambda = \frac{\sigma_y^2}{\sigma_x^2}$ the R_1^2 is given by:

$$R_1^2 = 1 - \frac{\Omega/\sigma_x^2}{\Omega/\sigma_x^2 + 1} \qquad \frac{\Omega}{\sigma_x^2} = -\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \frac{\lambda}{c}}.$$

Thus, R_1^2 is again falling in λ – that is, it is decreasing in the relative smoothness of the trend. In turn, priors over the smoothness of trends are used in the literature (e.g., Del Negro et al. (2017)).

5 Empirical applications

5.1 Natural rate of interest

One of the most pressing concerns of policymakers around the world has been for decades to infer the path the natural rate of interest. Several models have been developed to meet that need and most of these models have been criticized for being poorly identified. In this section, we use the methods developed earlier in the paper to evaluate the models developed by Holston, Laubach and Williams (2023) and Bauer and Rudebusch (2020) to estimate the natural rate of interest.

Table 1 shows the metrics we developed in the previous section to compare the performance of the two models in estimating r^*. The significant gap in performance is remarkable and easy to spot at a first

glance, as the model by Bauer and Rudebusch (BR) is characterized by lower uncertainty, higher R^2 and lower half-life.[9] Subsections 4.1 and 4.2 provide a brief description of the models and then dig deeper into the statistics presented in Table 1.

Table 1: Summarizing model performance and measures for i^* (BR) and r^* (HLW)

	BR	HLW: levels	HLW: 5-yr smoother	HLW: 5-yr change
Median s.d. of state estimate given parameters	0.18pp	1.5pp	1.4pp	0.8pp
Average s.d. of state estimate	0.84pp	2.2pp	2.0pp	1.0pp
Median 4-quarter ahead R^2	0.98	0.07	0.02	0.08
Median 20-quarter ahead R^2	0.99	0.28	0.20	0.16
Median half-life of estimation errors [quarters]	1	56	80	7

5.1.1 Bauer and Rudebusch (2020)

The model assumes that nominal government yields share three common factors; each common factor can be decomposed in an idiosyncratic cyclical component and a shared random walk. The shared random walk component represents the nominal rate of interest. The main equations of this model are as follows

$$Y_t^o = A_{BR} + B_{BR} P_t + e_t \tag{5.1}$$

$$\begin{aligned} P_t &= \bar{P} + \gamma \tau_t + \tilde{P}_t \\ \tilde{P}_t &= \Phi \tilde{P}_{t-1} + \tilde{u}_t \\ \tau_t &= \tau_{t-1} + \eta_t \end{aligned}$$

where Y_t^o is a vector of 17 yields, e_t is a vector of 17 uncorrelated measurement errors i.i.d. normal with zero mean and the same standard deviation σ_e, P_t is a vector of three unobserved factors, η_t is i.i.d. normal with standard deviation σ_η, $\tilde{u}_t$ is i.i.d. normal with covariance matrix $\tilde{\Omega}$. The sample period used for the estimation is 1971q1 to 2018q1. In Figure 1 and 4, we use PTR from the FRB/US to translate the nominal rate of interest τ into r^*. Figure 4 shows that r^* has increased from the mid 1970s to the early 2000s and it has declined since then.

[9] The large difference in standard deviations is particularly striking given that the BR estimate is in nominal terms and the LW estimate is in real terms.

Figure 4: Filtered estimate of τ - inflation expectations

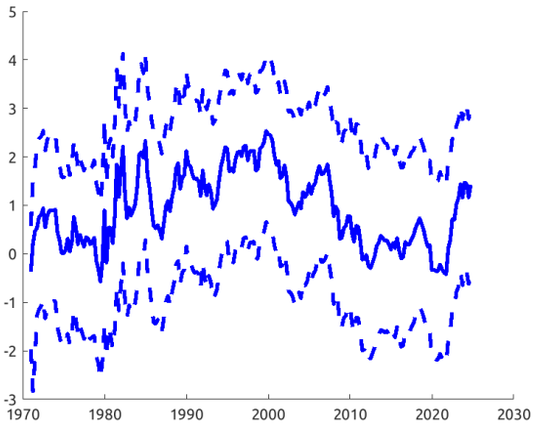
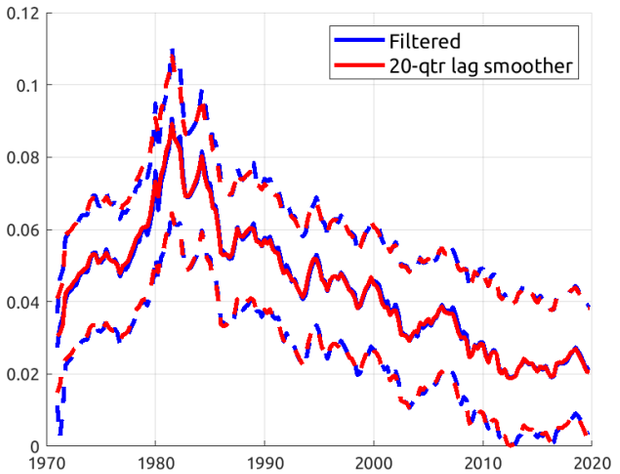


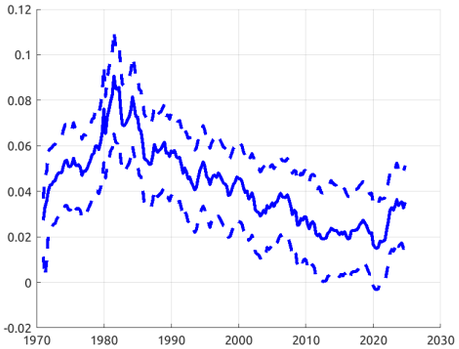
Figure 5: Filtered and smoothed estimates of τ_t: $\hat{\tau}_{t|t}$ and $\hat{\tau}_{t-20|t}$



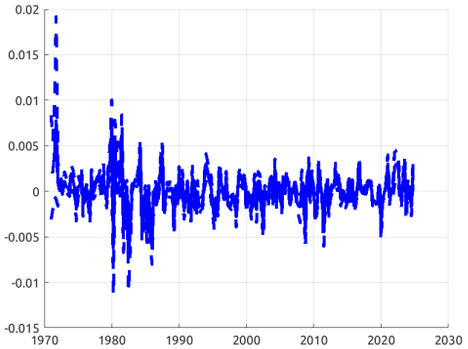
We believe that the first step in evaluating the performance of a model in uncovering a hidden state should be to look at the correlation between the true state and the estimated one. Figure 7 plots the histogram of the unconditional correlations for all draws of the shock $\eta = \tau_t - \tau_{t-1}$ and of the 20-period change in τ. The model does a pretty good job as the median correlations are 0.83 and 0.99, respectively. In Figure 8, we look at the correlation between the true τ and its estimated counterpart at specific horizons. Since τ is assumed to be a random walk, as discussed earlier, this horizon-specific correlation converges to one. The convergence happens quickly, as the median correlation is already 0.91 at horizon 1. The second step in the evaluation is to look at the half-life, shown in Figure 9. Not surprisingly, the model of Bauer and Rudebusch has very quick speed of convergence, with a median half life of just one quarter.

Figure 6: Filtered estimates of τ_t: level estimate, one-quarter change, and 20-quarter change

(a) Kalman filter



(b) 1-quarter change



(c) 20-quarter change

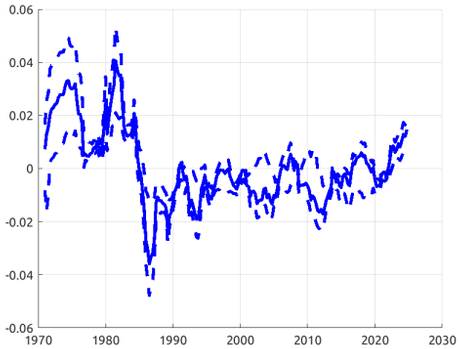
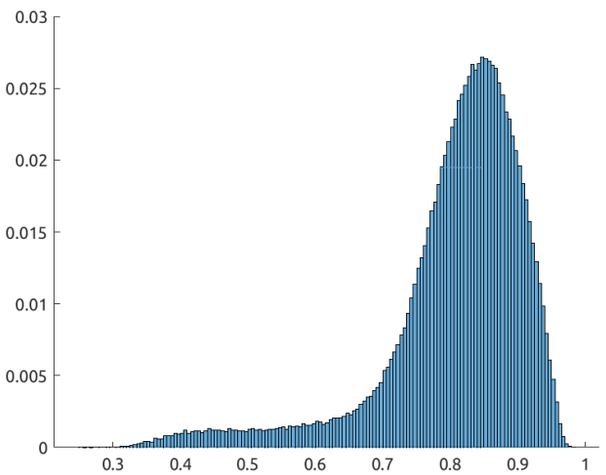


Figure 7: Unconditional correlations of 1-quarter and 20-quarter change in τ_t

1-quarter change



20-quarter change

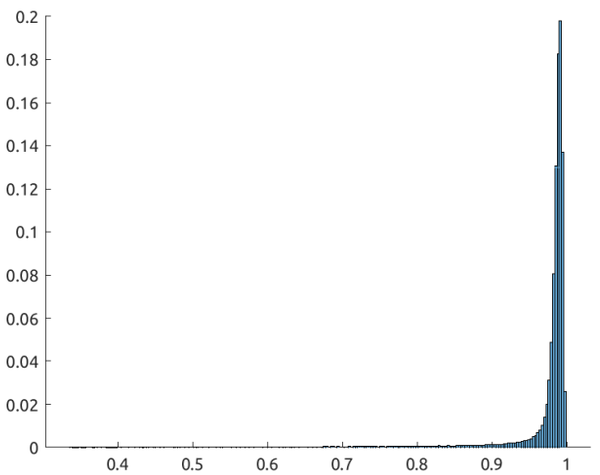


Figure 8: Correlations of τ: $\hat{\tau}_{t|t}$ at different horizons

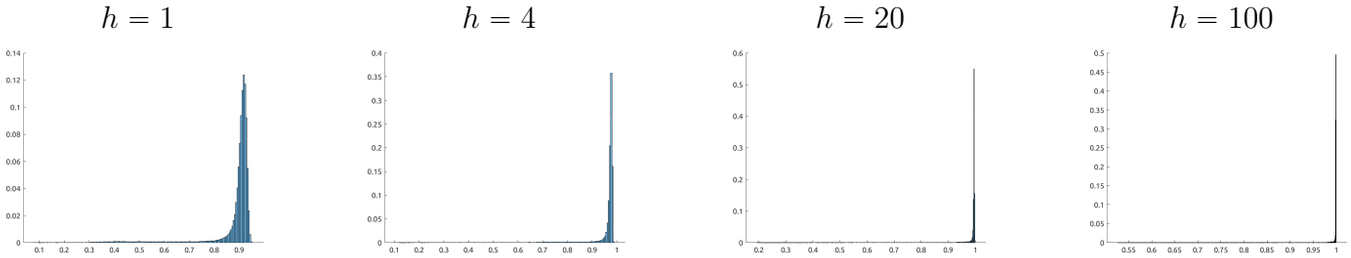
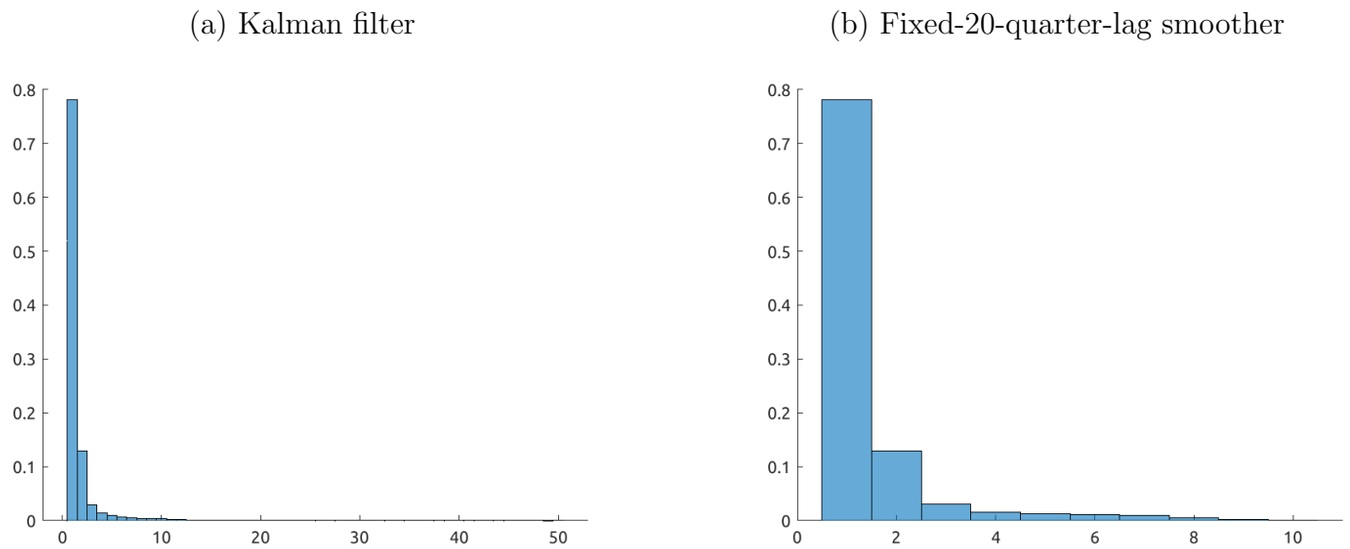


Figure 9: Half-life of τ_t estimation errors: $\hat{\tau}_{t|t}$ vs 20-quarter change



5.1.2 Holston, Laubach and Williams (2023)

We estimate the Holston, Laubach and Williams (2023) version of the original Laubach and Williams (2003) model using Bayesian methods. As in one of the variants of the Laubach and Williams (2003) model, both technology levels and real interest rates are non-stationary. The sample period is 1960:q1 to 2025:q2. Table 4 in the appendix lists the prior and posterior for the estimated parameters. We use data from the NY Fed website.[10]

The economy is described by the following IS equation (relating the output gap $\tilde{y}_t$ to its own lags as well as the real interest rate gap) and a Phillips Curve (relating inflation to the output gap):

$$\tilde{y}_t = a_1\tilde{y}_{t-1} + a_2\tilde{y}_{t-2} + \frac{a_r}{2}\left[(r_{t-1} - r^*_{t-1}) + (r_{t-2} - r^*_{t-2})\right] + \sigma_{t,\tilde{y}}\varepsilon_{\tilde{y},t} \tag{5.2}$$

$$\pi_t = b_\pi\pi_{t-1} + (1 - b_\pi)\frac{\pi_{t-2} + \pi_{t-3} + \pi_{t-4}}{3} + +b_y\tilde{y}_{t-1} + \sigma_{t,\pi}\varepsilon_{\pi,t}. \tag{5.3}$$

[10] https://www.newyorkfed.org/research/policy/rstar accessed 10/31/2025.

The main variable of interest is r_t^*. Imposing a unitary elasticity of substitution, it evolves as the sum of the growth rate g_t and the interest-rate specific component z_t:

$$r_t^* = c \cdot g_t + z_t.$$

The other exogenous variables evolve as follows:

$$z_t = z_{t-1} + \sigma_z \varepsilon_{z,t} \tag{5.4}$$
$$y_t^* = y_{t-1}^* + g_{t-1} - \phi \cdot Covid_t + \sigma_{y*} \varepsilon_{y*,t} \tag{5.5}$$
$$g_t = g_{t-1} + \sigma_g \varepsilon_{g,t} \tag{5.6}$$
$$r_t = r_{t-1} + \sigma_r \varepsilon_{r,t}, \tag{5.7}$$

where $Covid_t$ is the Oxford COVID-19 Government Response Stringency Index, which is intended to capture the temporary negative supply effects of the pandemic. Note that the intensely turbulent data developments during the pandemic are accommodated by allowing for heteroskedastic shocks in the IS and PC equations in the 2020-2022 period. [11]

We have access to the following measurements: r_t, π_t, output growth Δy_t^{obs}, and $Covid_t$, where:

$$y_t = 100 \cdot log(GDP) = y_t^* + \tilde{y}_t \tag{5.8}$$
$$\Delta y_t^{obs} = (y_t - y_{t-1}). \tag{5.9}$$

While the model has implications for both the output gap and the natural rate of interest r_t^*, we focus on the latter. We begin by presenting the level estimates of r_t^* in Figure **??** – which we previewed in Figure 1 in the introduction.

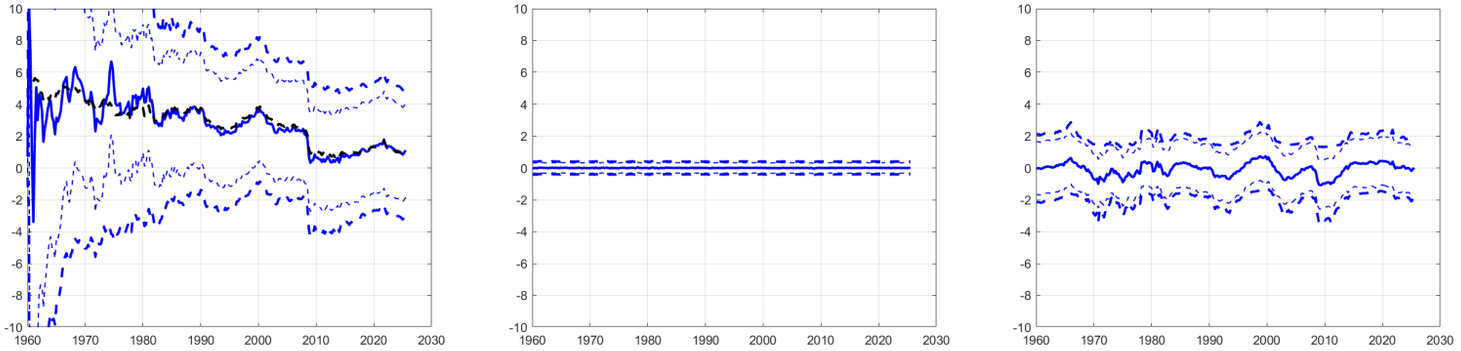
As in the introduction, Figure 10(a) reveals first and foremost a large degree of uncertainty. This can be seen from the dashed lines, which show the 95% and 90% uncertainty bands. For the filter, shown in blue, the band is around 10pp wide once the uncertainty from the diffuse initialization of the Kalman filter has settled. The fixed-lag smoother, than uses 20 quarters of future data, has an uncertainty band that is only slightly narrower (not shwon). Most of the uncertainty is state uncertainty: The median state uncertainty is 1.5pp for the Kalman filter and 1.4pp for the Kalman smoother – about 70% of the average standard deviation of the state estimate of 2.2pp and 2.0pp, respectively. Given parameters, the smoother thus reduces the uncertainty only little.

Given the large degree of uncertainty, practitioners may be hesitant to use the level estimates of this version of the Holston, Laubach and Williams (2023) model. Can the model be useful about changes? Figure 10 (b) and (c) show the evolution and estimation uncertainty of the Kalman filter estimates of the 1-quarter-change and the 4-quarter change. The uncertainty is much smaller. For the 1-quarter change,

[11] Heteroskesticity is introduced by scaling variance with year-specific estimated factors for 2020, 2021, and 2022.

Figure 10: Filtered estimates of r_t^*: level estimate, one-quarter change, and 20-quarter change

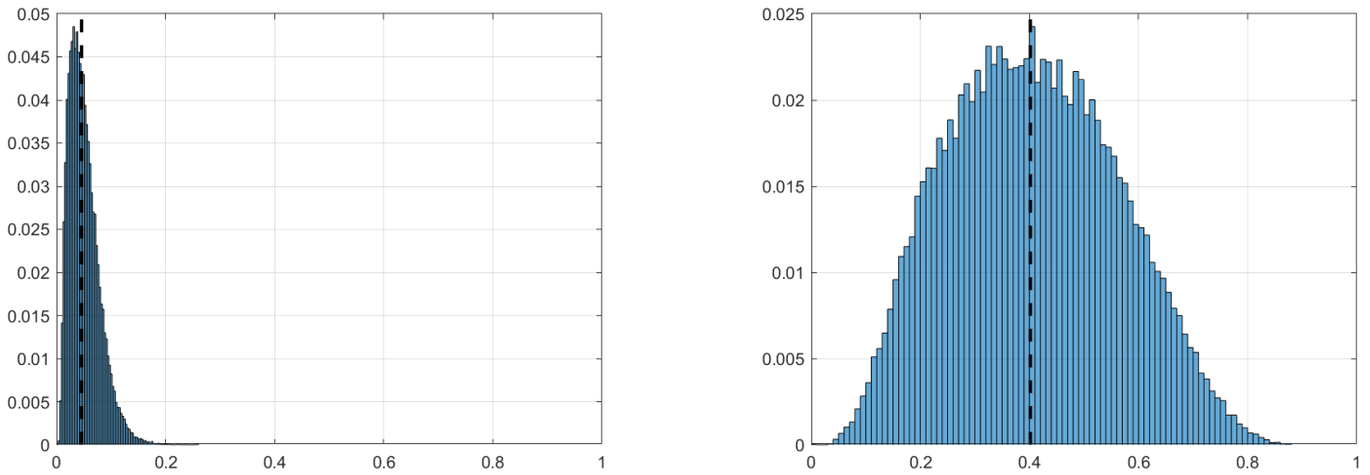
(a) Kalman filter (b) 1-quarter change (c) 20-quarter change



the 95% uncertainty band is about 0.8pp wide. However, quarter-to-quarter changes are irrelevant unless they accumulate. Thus, we also analyze the 20-quarter change. Here, the uncertainty band is about 5pp wide. This is still less than half of the uncertainty about the level of the state. Moreover, the mean estimate shows some persistent patterns, such as the sudden drop in r_t^* early during the Great Recession.

Figure 11: Unconditional correlations of 1-quarter and 20-quarter change in r_t^*: $\hat{r}_{t|t}^*$ at different horizons

1-quarter change 20-quarter change

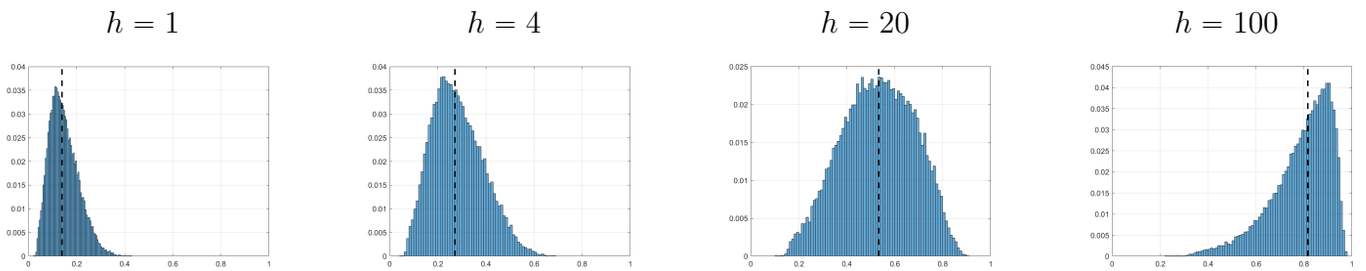


Are the patterns of the filtered changes predictive of actual changes? Figure 11 shows histograms the correlation of the filtered changes with the actual changes. Changes are stationary, so we can look at the unconditional correlation. There is substantial parameter uncertainty about these correlations. For the 1-quarter change, the median is about 0.05, implying a median R^2 around .25%. The correlation with the true change is larger for the 20-quarter change, implying a median R^2 of 10%.

How closely does the – very uncertain – level estimate vary with the truth? Since the level is non-

statonary but cointegrated with the filtered estimate, Theorem 3.2 suggests that the long-run correlation is meaningless: Cointegrated variables comove in the long-run. But how quickly does the system reach this long-run? Figure 12 shows the conditional correlation for various horizons.

Figure 12: Conditional correlations of r_t^*: $\widehat{r}_{t|t}^*$ at different horizons



The conditional correlations of r_t^* with its mean estimate start out low, even lower than unconditional correlation for 1-quarter changes. The median correlation is around 0.1 and the R^2 is thus 1%. At the 4-quarter horizon this R^2 has risen to about 4%. At 20 quarters, the median R^2 is about 20%. Overall, the Holston, Laubach and Williams (2023)-estimates converge much more slowly than the Bauer and Rudebusch (2020) estimates. This makes the estimates difficult to use at short- to medium-run frequences. However, as we discuss below, the model is more reliable for the output gap.

Figure 13: Half-life of r_t^* estimation errors: $\hat{r}_{t|t}^*$ vs 20-quarter change

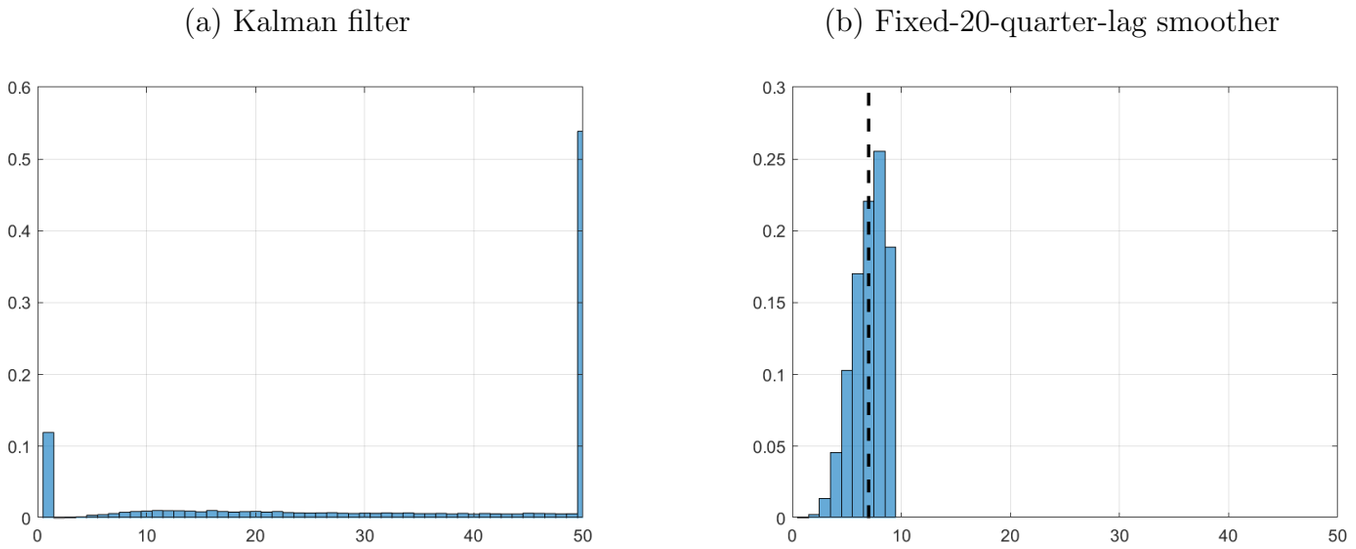


Table 2 summarizes these measures of uncertainty for r^* and compares the performance to the analogous measures for the output gap. The output gap has much higher reliability measures, even though its state uncertainty is of comparable magnitude – relative to the standard deviation of the innocation to the state – than that for r^*.

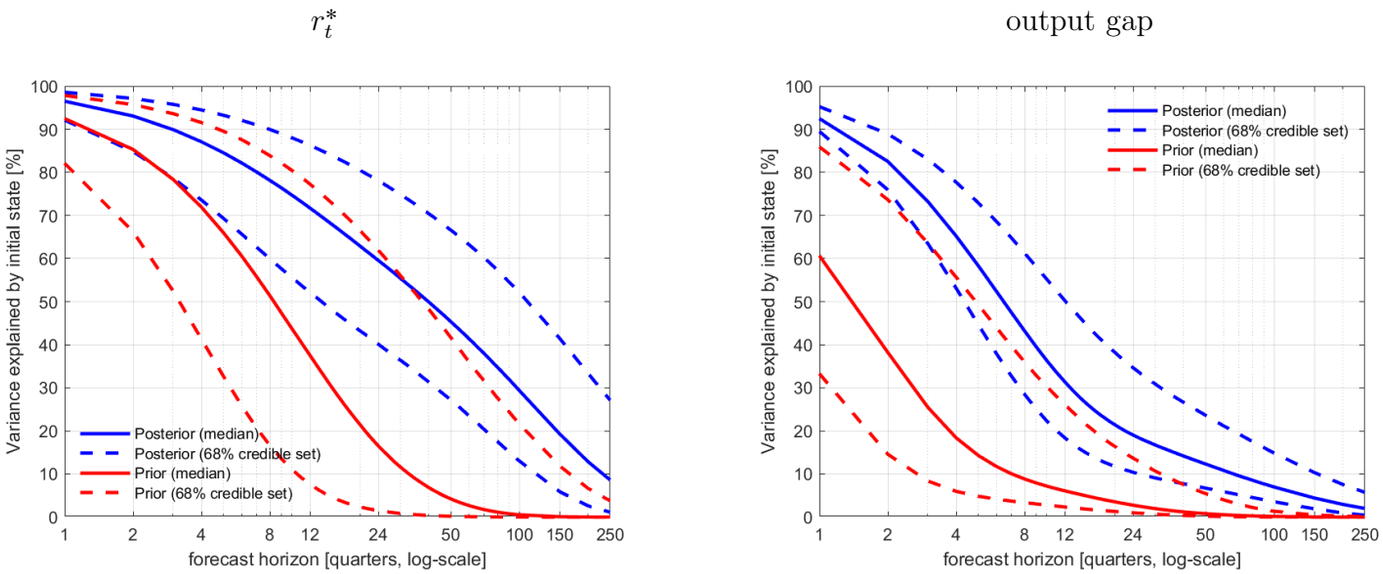
Turning to dissecting the source of the estimation uncertainty, Figure 13 shows the distribution of half-lives of the filtering errors for the level estimates (left panel) and the 20-quarter changes. For the level

Table 2: Summarizing model performance and measures for y^* and r^* in the HLW model

	y-gap	r^*	r^*: 5-yr smoother	1-qtr r^*-change	5-yr r^*-change
Median s.d. of state estimate					
given parameters/innovation s.d.	3.8pp	7.6pp	6.9pp	0.8pp	3.9pp
Average s.d. of state estimate/innovation s.d.	5.7pp	10.8pp	10.0pp	1.0pp	4.9pp
Innovation s.d. (post mean)	0.5pp	0.2pp	0.2pp	0.2pp	0.2pp
Median 4-quarter ahead R^2	0.27	0.07	0.02	0.00	0.08
Median 20-quarter ahead R^2	0.82	0.28	0.20	0.00	0.16
Median half-life of estimation errors [quarters]	10	56	80	1	7

estimate, the half-life across parameter draws ranges from 1 period to more than 50 [the plot is truncated at 50 quarters]. The median half-life is 56 quarters. On average, errors thus take 14 years to be reduced by half. For long-changes, this half-life is much better: The median is of only 7 quarters. However, changes have only a low R^2.

Figure 14 shows the FEV contribution of the initial state for $r*$ (left panel) and the output gap (right panel). The initial state matters more for r^* than for the output gap: The posterior median implies that it takes more than 12 years for the contribution to fall below 50%, as opposed to two and a half years for the output gap (solid blue lines). These posterior estimates are much higher than the prior beliefs, shown in red.



Plotted is $100 - 100 \times \frac{\left[\sum_{h=1}^{H}(A-KC)^{h-1}(B-KD)(B-KD)'((A-KC)')^{h-1}\right]_{ii}}{[\Omega_{t|t}]_{ii}}$ against H.

Figure 14: Contribution of the initial to the state estimation uncertainty by horizon and state of interest

Given that the posterior gives a more important role to the initial state than the prior, one may expect

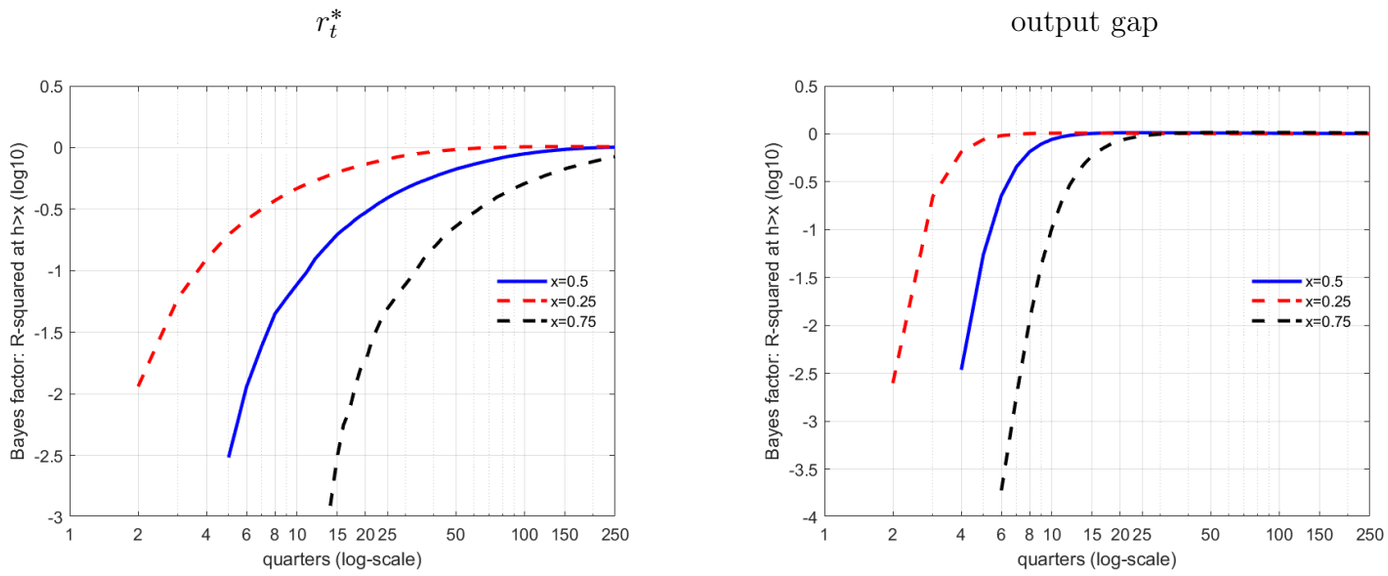


Figure 15: Posterior vs prior probability of R-squared of the KF with the truth exceeding 0.5 or 0.25 at different horizons

the data overall to favor less reliability. This is indeed the case. Figure 15 shows the Bayes Factor in the form of the $\log_{10}$ posterior-to-prior probability of restrictions on R_h^2 for r^* and the output gap against the horizon h at which the conditional R_h^2 is restricted. Imposing a given threshold at a shorter horizon is less favored by the data. Requiring an $R_h^2 \geqslant 0.5$ at horizon 8 for r^*, for example, implies a Bayes factor of $10^{-1.5}$ – overwhelming evidence against that restriction. The evidence against these restrictions weakens only slowly with the horizon. In contrast, the data provide little evidence against the same restriction for the output gap at horizon 8.[12]

Overall, the Holston, Laubach and Williams (2023) model has little power to recover r^* in the short- to medium-run. It does better for the output gap.

Del Negro et al. (2017) have shown that using longer-term yields and data on expectations can yield precise estimate of r^*. Similarly, Kim and Wright (2005) show how to incorporate expectations of future yields in a term structure model. In ongoing work, we use SPF stock return expectations as a predictor. Absent risk-premia, returns reflect investors' marginal utility – which should be driven by expected growth and other determinants of the desire so save. Thus:

$$\mathbb{E}_t^{SPF}[R_{t\to t+10\text{years}}] = \alpha + r_t^* + \nu_t,$$

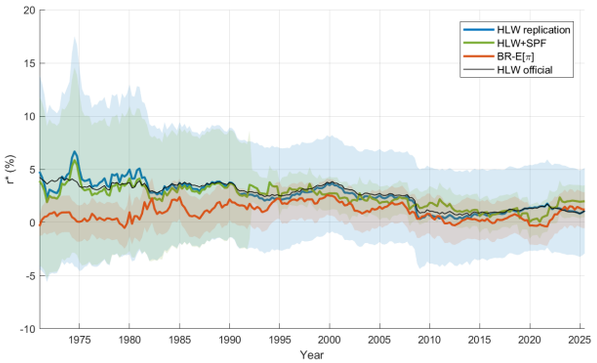
where we assumed that r_t^* is *iid*. α captures an average risk premium.[13]

[12]Figure 23 shows the corresponding conditional tracking estimates. The restrictions can affect both the central tendency and the width of the posterior uncertainty.

[13]We address missing data in the SPF – the long-term expectations are elicited only once a year – by interpolating the data using the most recent survey and yields. We treat the interpolated observations as having higher variance. When

Introducing return expectations into the HLW model yields point estimates that are overall consistent with the baseline HLW model. However, in recent years the expectations-augmented model implies a a noticeable increase $\hat{r}_t^*$ in the aftermath of the pandemic. See Figure 16.

With 95% error bands



Point estimates

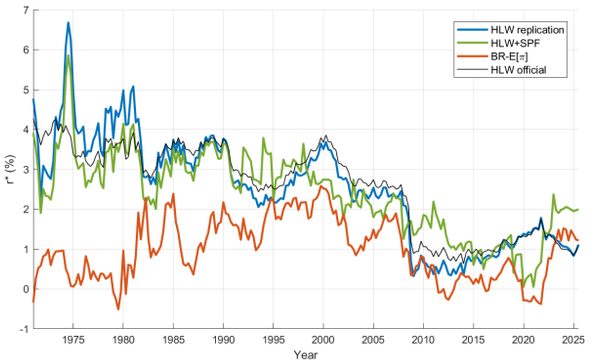


Figure 16: Incorporating return expectations in the HLW model

5.2 A simple trend-cycle model of the economy

In this section, we evaluate a standard unobserved-components model with output, inflation, and unemployment:

$$y_t^{\text{obs}} = y_t^* + c_t + \sigma_y \eta_t^y \tag{5.10}$$

$$\pi_t^{\text{obs}} = \pi_t^* + b_1 c_t + b_2 c_{t-1} + \sigma_\pi \eta_t^\pi \quad \text{(Phillips curve)} \tag{5.11}$$

$$u_t^{\text{obs}} = u_t^* + c_1 c_t + c_2 c_{t-1} + \sigma_u \eta_t^u \quad \text{(Okun's law)} \tag{5.12}$$

where y is the logarithm of GDP, π is inflation, u is the unemployment rate, y^* is potential output, c is the output gap, π^* is trend inflation, and u^* is the natural rate of unemployment. The trends (y^*, π^*, u^*) follow random walks and the output gap follows an AR(2) process.

Table 3 summarizes the reliability measures. While posterior standard deviations are of similar magnitude across states, reliability differs markedly. Potential output and trend inflation are estimated with high reliability: their median R^2 exceeds 0.85 at a four-quarter horizon and approaches one at longer horizons. This likely reflects their direct link to observables through GDP and inflation.

In contrast, the natural rate of unemployment u_t^* is estimated much less precisely. Its median four-quarter R^2 is only 0.04, rising to 0.16 even at a twenty-quarter horizon. Despite a posterior standard deviation that is not dramatically larger than for the other states, the data contain little information that allows the filter to disentangle the natural rate from cyclical movements in unemployment. In this

experimenting with removing the effect of risk premia by projecting out term premia, the results were largely unchanged.

Table 3: Summarizing model performance

	c	u^*	y^*	π^*
Median s.d. of state estimate given parameters*	14.4pp	28.9pp	17.2pp	6.9pp
Average s.d. of state estimate*	16.5pp	31.1pp	19.7pp	7.7pp
Innovation s.d. (post mean)	0.37	0.05	0.97	0.59
Median 4-quarter ahead R^2	0.69	0.04	0.85	0.89
Median 20-quarter ahead R^2	0.87	0.16	0.97	0.97
Median half-life of estimation errors [quarters]	77	81	72	1

Notes: *The first two rows are standardized by the estimated standard deviation of the innovation to each state equation (Table 5).

model, unemployment depends on the output gap through Okun's law, implying that information about u_t^* must be inferred only indirectly through other latent variables. As a result, the Kalman filter struggles to distinguish persistent shifts in the natural rate from cyclical fluctuations.

The output gap c lies between these extremes. Its median four-quarter R^2 of 0.69 suggests that the model can recover a substantial share of cyclical variation in real time, and reliability increases further at longer horizons. Nevertheless, the gap still exhibits considerable uncertainty relative to potential output and trend inflation, reflecting the difficulty of distinguishing cyclical movements in activity from changes in potential output.

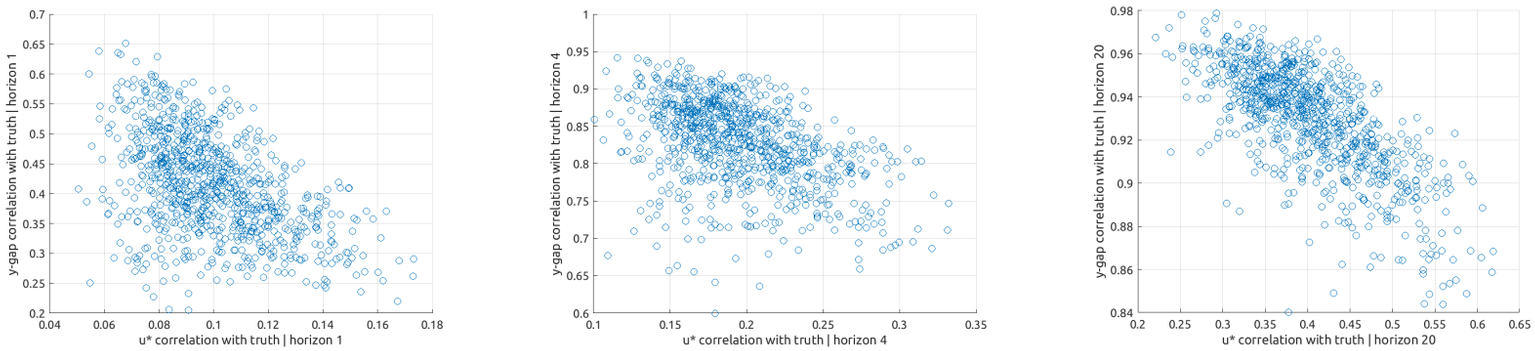
The half-life statistics further clarify these differences. Estimation errors for the gap, the natural rate of unemployment, and potential output are highly persistent, with median half-lives on the order of 70–80 quarters. This implies that uncertainty about these states dissipates only slowly as new data arrive. In contrast, the half-life for trend inflation is only one quarter, indicating that estimation errors are corrected rapidly once additional inflation data become available.

Figure 17 provides a complementary perspective by comparing the horizon-specific correlations between the true states and their filtered estimates for the output gap c and the natural rate of unemployment u^*. While both variables are persistent, the figure shows that the output gap can be recovered substantially more reliably from the data than the natural rate of unemployment. The correlation between c and its filtered estimate rises quickly with the horizon and approaches values close to one as additional observations accumulate. In contrast, the correlation between u^* and its filtered estimate remains much lower even at longer horizons. Moreover, there seems to be evidence of a tradeoff between the reliability of c and the reliability of u, which could again reflect the difficulty of distinguishing cyclical movements from changes in the natural rate.

Taken together, these results illustrate an important point emphasized earlier in the paper: posterior uncertainty alone is not sufficient to assess the reliability of latent-state estimates. Variables with similar posterior standard deviations may differ markedly in how much of their variation can be recovered from

Figure 17: Correlation of $\hat{c}_t^*$ with the truth at various horizons vs correlation of $\hat{u^*}_t^*$ with the truth (posterior)

$h = 1$ $h = 4$ $h = 20$



the data. The reliability measures developed in Section 3 therefore provide additional insight into which latent variables can be estimated credibly and which remain weakly identified in practice.

6 Assessing models

6.1 Assessing predictions for observables

The reliability measures developed in the previous sections quantify how well a state-space model can recover latent variables such as the natural rate of interest, potential output, or the output gap. These measures depend on the parameters of the model and on the signal-to-noise structure implied by the measurement and transition equations. However, an important limitation remains: because the states are unobserved, we cannot directly verify whether the predictive distribution of the model accurately characterizes the uncertainty surrounding the latent variables.

In particular, a large posterior variance of a latent state estimate is not necessarily problematic. High uncertainty may simply reflect the fact that the data contain limited information about that state. Conversely, a model may produce a narrow posterior distribution while still mischaracterizing the true dynamics of the underlying economy. Since the latent states themselves are not observable, it is impossible to directly evaluate whether the estimated uncertainty bands around these states are correctly calibrated.

However, the predictions for both observables and unobservables are tied together by forecasts for the model's state vector: The h-step ahead forecast for the observables is $CA^{h-1}\hat{x}_{t|t}$ and the h-step ahead forecast error variance is $C\,\mathrm{Var}[x_{t+h-1}|t]C' + DD'$. We therefore propose complementing the reliability analysis with tests based on observable variables. The key idea is that even though the latent states cannot be observed their model-based forecasts imply predictions for observable variables. If the model correctly characterizes the dynamics of the economy and the uncertainty surrounding the latent states, these predictions should be consistent with the empirical behavior of the observables.

A simple starting point is to evaluate the one-step-ahead forecast errors of the observables,

$$y_t - C\hat{x}_{t-1|t-1},$$

which correspond to the innovations of the Kalman filter. If the model is correctly specified, these forecast errors should behave according to the distribution implied by the model.

A more stringent test examines the predictive performance at longer horizons. Let the h-step-ahead forecast error for observable y be

$$y_{t+h} - CA^{h-1}\hat{x}_{t|t}.$$

These forecast errors capture the model's ability to predict future observable outcomes based on current information. If the state-space model correctly describes the dynamics of the economy, the statistical properties of these forecast errors should be consistent with those implied by the model.

In analogy to the infeasible regression (3.3), we can quantify the reliability of the model for the observables using the following regression:

$$y_{i,t+h+1} - \hat{y}_{i,t+h+1|t} = \alpha + \beta\big(\hat{y}_{i,t+h+1|t+h} - \hat{y}_{i,t+h+1|t}\big).$$

The right-hand side measures the revision in the forecast of the observable between time t and time $t+h$. Intuitively, if the model uses information efficiently, these forecast revisions should explain the unexpected changes in the observable variable. The R^2 of this regression therefore provides a natural measure of how well the model translates new information into improved predictions for the observables.

Because the observables are directly measured in the data, we can compute the empirical counterpart of this R^2 statistic and compare it to the value implied by the model – unlike the reliability measure for the latent variables. Denote these quantities by

$$R^2_{y_i,h|\text{data}} \quad \text{and} \quad R^2_{y_i,h|\text{theory}},$$

respectively. The difference

$$\Delta R^2_{i,h} \equiv R^2_{y_i,h|\text{data}} - R^2_{y_i,h|\text{theory}} \tag{6.1}$$

provides a diagnostic for model misspecification.

In addition to the R^2 statistics, we also examine the root mean squared forecast error (RMSE) of the model. Together, the RMSE and the ΔR^2 statistics summarize how well the model predicts the observable variables at different horizons.

Evaluating the statistical significance of these statistics is not straightforward because they depend on filtered estimates of the latent states rather than directly observed quantities. As emphasized by Pagan and Robinson (2022), the distribution of such statistics can therefore be complicated.

To address this issue, we adopt a simulation-based approach. First, we compute the statistics of

interest (such as $\text{RMSE}_{i,h}$ and $\Delta R^2_{i,h}$) using the actual data and the estimated model parameters. Next, we simulate artificial datasets from the estimated model, typically evaluated at the posterior mean of the parameters. For each simulated dataset, we recompute the same statistics. This procedure yields an empirical distribution of the statistics under the null hypothesis that the model is correctly specified.[14]

Comparing the statistics computed from the actual data with the simulated distribution allows us to assess whether the model provides an adequate description of the observable dynamics of the economy. If the empirical statistic lies well within the simulated distribution, the model's predictions are consistent with the data. Conversely, if the empirical statistic falls far outside the simulated range, this provides evidence that the model fails to capture important aspects of the data-generating process.

Application to HLW. In our empirical application, we focus on the RMSE and the ΔR^2 statistics for inflation and output growth in the HLW model. These are the two endogenous observables in the model, while the real interest rate and the COVID stringency index are treated as exogenous variables.

Figure 18 compares the empirical R^2 statistic to the simulated one. The left panel shows inflation and the right panel output growth from one to 40 quarters ahead. The statistics are computed using both the time-varying Kalman Filter (shown in blue) and the steady state Kalman filter (orange). We can clearly reject the steady state Kalman filter, because the empirical estimates – shown as the orange line – lies outside the 90% band. However, for output growth with either filter and for the time-varying Kalman filter used in practice, we cannot reject that the model is correctly specified.

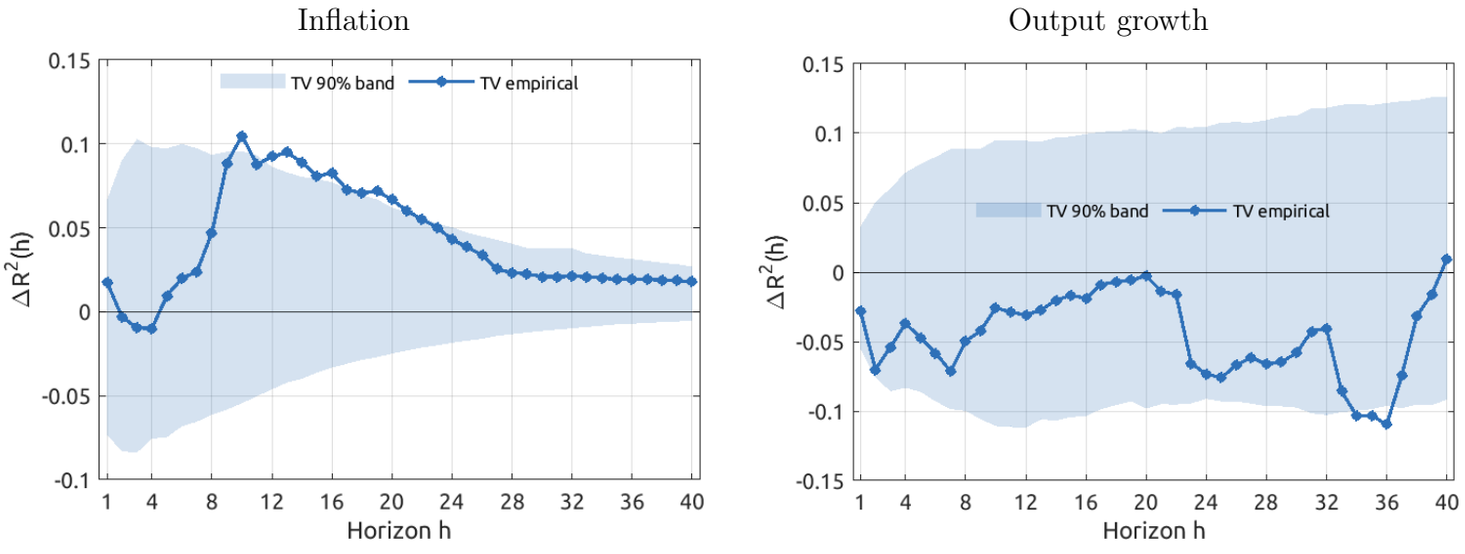


Figure 18: Gap between the theoretical and empirical R^2 in (6.1) for inflation and output growth at various horizons: Posterior mean vs simulated distribution in the HLW model.

[14]It is straightforward to incorporate parameter uncertainty in a Bayesian setting by computing the posterior predictive density. In this case, we re-estimate the posterior mean for each artificial dataset via importance sampling.

Turning to forecast performance in Figure 19, the simulated RMSE for inflation grows faster than the empirical estimates. However, also the empirical RMSE grows steadily with the horizon length, consistent with a unit-root in inflation. The gap between the model and the empirical estimates thus suggest that the estimated variance of the inflation shocks or other shocks in the model is too large. In contrast, the empirical output growth RMSE behaves as predicted by the model. In Appendix E, the interested reader can find charts showing ΔR_y^2 and RMSEs for the UC3 model.

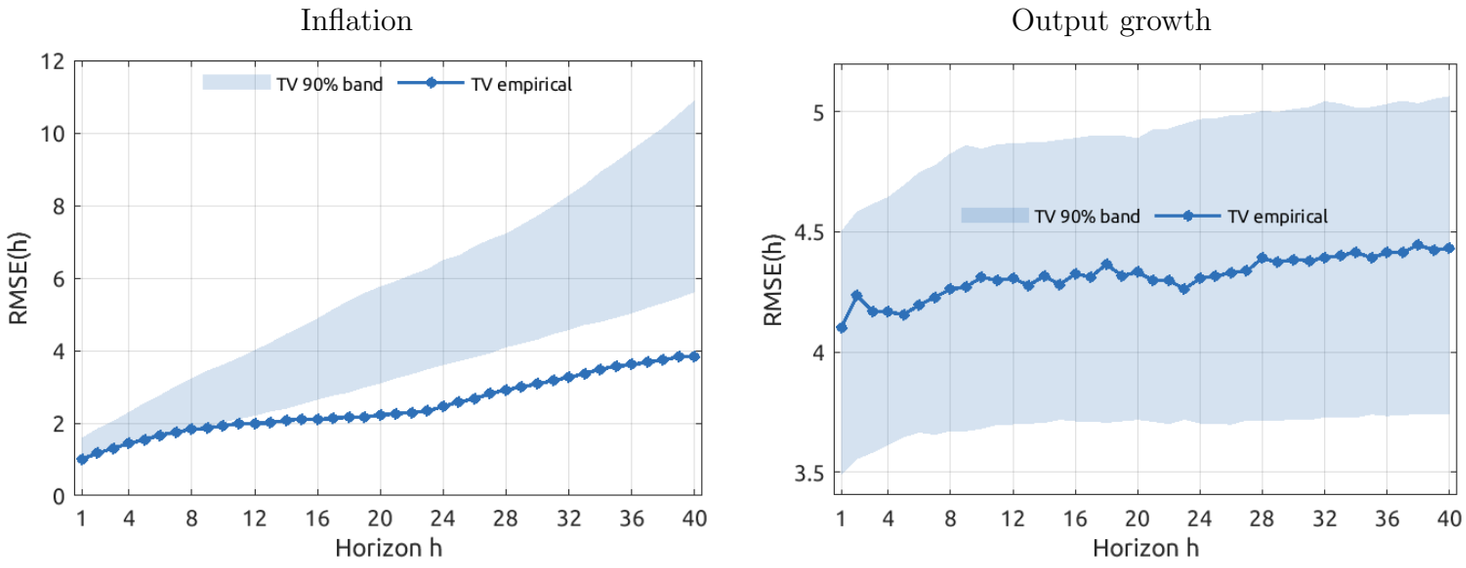


Figure 19: RMSE for inflation and output growth at various horizons: Posterior mean vs simulated distribution in the HLW model.

6.2 Implied restrictions on latent state reliability measures

A useful way to sharpen the interpretation of these observable diagnostics is to exploit the structural relationship between reliability measures for observables and latent states. It can be shown that the conditional reliability measure for a latent state, $R^2(x_i, h)$, is proportional to the corresponding measure for an observable y_j, with the proportionality factor given by a model-implied wedge that depends only on the state-space parameters and the horizon h. This implies that any discrepancy between model-implied and empirical reliability in observables,

$$\Delta R^2(y_j, h) \equiv R^2(y_j, h)^{\text{theory}} - R^2(y_j, h)^{\text{empirical}},$$

can be mapped into a corresponding discrepancy for the latent state,

$$\Delta R^2(x_i, h) = f(x_i, y_j, h)\, \Delta R^2(y_j, h),$$

where $f(x_i, y_j, h)$ is a function of model parameters. Importantly, $\Delta R^2(y_j, h)$ is directly observable from forecast-error and revision regressions, and therefore provides a testable measure of model misspecification. Under the maintained assumption that the state transition and measurement equations are correctly specified, the wedge $f(x_i, y_j, h)$ is invariant to these empirical deviations and governs how observable misspecification translates into latent-state reliability. This mapping provides a disciplined way to assess latent-variable reliability indirectly: while $\Delta R^2(x_i, h)$ itself is not observable, the model-implied transformation of $\Delta R^2(y_j, h)$ delivers an estimate of the corresponding reliability gap for the latent state. Moreover, the magnitude of $f(x_i, y_j, h)$ determines whether observable discrepancies are amplified or dampened in the latent state, reflecting the relative signal-to-noise ratios and alignment of the state and measurement equations. In this sense, evaluating $\Delta R^2(y_j, h)$ is not only a validation exercise for observables, but also an informative diagnostic for the credibility of the model's implied reliability of latent variables.

7 Conclusion

This paper proposes statistics to assess the reliability of state estimates in linear state-space models. Kalman filter estimates can be imprecisely estimated – and yet the data flow can be informative about changes in the target variable over several quarters. But estimation errors can also persist beyond business cycle frequencies, as evidenced by the implied half-life of estimation errors.

In our application, the Bauer and Rudebusch (2020) model dominates the Holston, Laubach and Williams (2023) model in terms of estimation uncertainty, predictive power, and half-life. However, it punts on the important question of the level of inflation expectations.

Our pessimistic assessment of the Holston, Laubach and Williams (2023) model is based on model-based uncertainty and reliability assessments. When applied to observables, the model suggests that the reliability statistics for observables are broadly in line with the data, even though the uncertainty of longer-term inflation forecasts is overstated relative to the data.

We show theoretically that high reliability can go hand-in-hand with high state uncertainty. This is not the case in the HLW model for r^* estimates. However, in a simple trend-cycle model of the U.S. economy we verify also empirically that reliability does not mechanically move with state uncertainty.

References

Bauer, M. D., and G. D. Rudebusch (2020): "Interest rates under falling stars," *American Economic Review*, 110(5), 1316–1354.

Buncic, D., A. Pagan, and T. Robinson (2023): "Recovering stars in macroeconomics," *Available at SSRN 4562801*.

Chahrour, R., and K. Jurado (2022): "Recoverability and expectations-driven fluctuations," *The Review of economic studies*, 89(1), 214–239.

Del Negro, M., D. Giannone, M. P. Giannoni, and A. Tambalotti (2017): "Safety, Liquidity, and the Natural Rate of Interest," *Brookings Papers on Economic Activity*, 48(1 (Spring), 235–316.

Diebold, F. X., and L. Kilian (2001): "Measuring predictability: theory and macroeconomic applications," *Journal of Applied Econometrics*, 16(6), 657–669.

Engle, R. F., and B. S. Yoo (1987): "Forecasting and testing in cointegrated systems," *Journal of Econometrics*, 35(1), 143–159.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97(3), 1021–1026.

Forni, M., L. Gambetti, and L. Sala (2019): "Structural VARs and noninvertible macroeconomic models," *Journal of Applied Econometrics*, 34(2), 221–246.

Hamilton, J. D. (2018): "Why You Should Never Use the Hodrick-Prescott Filter," *The Review of Economics and Statistics*, 100(5), 831–843.

Harvey, A. (1990): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hodrick, R. J., and E. C. Prescott (1997): "Postwar US business cycles: An empirical investigation," *Journal of Money, Credit and Banking*, 29(1), 1–16.

Holston, K., T. Laubach, and J. C. Williams (2023): "Measuring the Natural Rate of Interest after COVID-19," Staff Report 1063, Federal Reserve Bank of New York.

Kim, D. H., and J. H. Wright (2005): "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates," Finance and Economics Discussion Series 2005-33, Board of Governors of the Federal Reserve System (U.S.).

Laubach, T., and J. C. Williams (2003): "Measuring the natural rate of interest," *Review of Economics and Statistics*, 85(4), 1063–1070.

Pagan, A., and T. Robinson (2022): "Excess shocks can limit the economic interpretation," *European Economic Review*, 145, 104120.

Plagborg-Møller, M., and C. K. Wolf (2022): "Instrumental Variable Identification of Dynamic Variance Decompositions," *Journal of Political Economy*, 130(8), 2164–2202.

Sims, C. A., and T. Zha (2006): "Does monetary policy generate recessions?," *Macroeconomic Dynamics*, 10(2), 231–272.

A Proof of results on unit-root systems

Definition 4. *The pair (A, B) is completely controllable if and of the following equivalent conditions hold:*

1. *range*$[B, AB, \ldots, A^{n-1}B] \supset$ *range*A^n.
2. $w'A^iB = 0$ *for* $i = 0, 1, \ldots, n-1$ *implies* $w'A^n = 0$.
3. $w'B = 0$ *and* $w'A = \lambda w'$ *implies* $\lambda = 0$ *or* $w = 0$.

Definition 5. *The pair (A, B) is completely stabilizable if any of the following equivalent conditions holds:*

1. $w'B = 0$ *and* $w'A = \lambda w'$ *for some constant* λ *implies* $|\lambda| < 1$ *or* $w = 0$.
2. *There exists an* $n \times r$ *matrix* K *such that* $\lambda_i(A + BK')| < 1$ *for all* i.
3. *If (there exists a nonsingular* T*)* $TAT^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, $TB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$ *with* (A_{11}, B_1) *completely controllable, then* $|\lambda_i(A_{22})| < 1$.

Definition 6. *The pair (A, B) is completely detectable as (A', B) is completely stabilizable.*

Theorem 7. *(Anderson & Moore, 1979, p. 77) If the signal process model* $x_{k+1} = Fx_k + Gw_k$, $z_k = H'x_k + v_k$ *(*$\mathbb{E}[ww'] = Q, \mathbb{E}[vv'] = R, \mathbb{E}[wv'] = 0$*) is time invariant and not necessarily asymptotically stable, but the pair* (F, H) *is completely detectable and the pair* (F, GG_1) *is completely stabilizable for any* G_1 *with* $G_1G_1' = Q$*, then (a) for any nonnegative symmetric initial condition* $\Sigma_{k_0|k_0-1}$ *one has* $\lim_{k\to\infty} \Sigma_{k+1|k} = \bar{\Sigma}$*, with* $\bar{\Sigma}$ *independent of* $\Sigma_{k_0|k_0-1}$ *and satisfying* $\bar{\Sigma} = F(\bar{\Sigma} - \bar{\Sigma}H(H'\bar{\Sigma}H + R)^{-1}H'\bar{\Sigma})F' + GQG'$ *and (b)* $|\lambda_i(F - K_{AM}H')| < 1$ *where* $K_{AM} = F\bar{\Sigma}H(H'\bar{\Sigma}H + R)^{-1}$.

We can now prove Theorem 1 from the paper:

Proof. **Step 1.** Let's guess and verify that the cointegrating matrix is $[I, -I]$. Then $x_t - \hat{x}_{t|t} = Ax_{t-1} + B\epsilon_t - (KCx_{t-1} + (A-KC)\hat{x}_{t-1|t-1} + KD\epsilon_t) = (A-KC)(x_{t-1} - \hat{x}_{t-1|t-1}) + B\epsilon_t - KD\epsilon_t$. Thus, if $(A-KC)$ is stable, then $Z_t = x_t - \hat{x}_{t|t}$ is stationary – proving cointegration. Here, $A - KC$ is the speed of adjustment.

Step 2. We show that the assumptions of the Theorem imply that Theorem 7 applies.

First, let's express out system in terms of the classic state space representation. Write the shock vector as $\epsilon_t = [\tilde{\epsilon}_t', v_t']'$ and the state vector as $\tilde{x}_{t+1} \equiv [x_t', \tilde{\epsilon}_{t+1}']'$, so that:

$$\begin{bmatrix} x_{t+1} \\ \tilde{\epsilon}_{t+2} \end{bmatrix} = \underbrace{\begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}}_{\equiv F} \begin{bmatrix} x_t \\ \tilde{\epsilon}_{t+1} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ I \end{bmatrix}}_{\equiv G} \tilde{\epsilon}_{t+2} \tag{A.1}$$

$$y_{t+1} = \underbrace{\begin{bmatrix} C & \tilde{D}_\epsilon \end{bmatrix}}_{\equiv H'} \begin{bmatrix} x_t \\ \tilde{\epsilon}_{t+1} \end{bmatrix} + \tilde{D}_v v_{t+1}, \tag{A.2}$$

where v_{t+1} represent measurement error, so that $D\epsilon_{t+1} = [\tilde{D}_\epsilon, \tilde{D}_v][\tilde{\epsilon}'_{t+1}, \tilde{v}'_{t+1}]'$. Note that $Q = I, R = I$, so that $G_1 = I$.

From Definition , detectability of (F, H) is equivalent to stabilizability of (F', H). By the Hautus-Lemma for stabilizability, we need that for the eigenvalues λ of F' with $|\lambda| \geqslant 1$ that $M \equiv \begin{bmatrix} F' - \lambda I & H \end{bmatrix} = \begin{bmatrix} A' - \lambda I & 0 & C' \\ B' & -\lambda I & \tilde{D}'_\epsilon \end{bmatrix}$ has full row rank. Since the eigenvalues zero and the eigenvalues of A, we can focus on the unstable eigenvalues of A ith $|\lambda| \geqslant 1$. Equivalently, we can check that the column rank of M' (as in the the Hautus-Lemma for detectability in discrete-time systems).

For the unstable eigenvalues of A we thus the check column rank of $M' = \begin{bmatrix} A - \lambda I & B \\ 0 & -\lambda I \\ C & \tilde{D}'_\epsilon \end{bmatrix}$.

Since $|\lambda| \geqslant 1$, $-\lambda I$ is invertible. The columns corresponding to the second block thus have full rank.

Since (A, C) is detectable, for any eigenvalue λ of A with $|\lambda| \geqslant 1$, the matrix $\begin{bmatrix} A - \lambda I \\ C \end{bmatrix}$ has full column rank, which implies $\begin{bmatrix} A - \lambda I \\ C' \end{bmatrix}$ has full row rank.

Therefore, for all eigenvalues of A with $|\lambda| \geqslant 1$, the test matrix has full column rank.

We need to verify that $(F, GG_1) = (F, G)$ is stabilizable. By the Hautus-Lemma for stabilizability, we need that for the eigenvalues λ of F with $|\lambda| \geqslant 1$ that $\begin{bmatrix} F - \lambda I & G \end{bmatrix} = \begin{bmatrix} A - \lambda I & B & 0 \\ 0 & -\lambda I & I \end{bmatrix}$ has full row rank. Again, we can focus on the unstable eigenvalues of A.

Since (A, B) is stabilizable we have that $\begin{bmatrix} A - \lambda I & B \end{bmatrix}$ has full row rank. Moreover, since I is full rank, also the right-most block has full rank. It follows that the conditions of Theorem 7 are satisfied.

Step 3. We show that the Anderson-Moore Kalman filter with Kalman gain K_{AM} is equivalent to our Kalman filter with the Kalman Gain K.

Since y_{t+1} has no signal about $\tilde{\epsilon}_{t+2}$, it can be verified that $\bar{\Sigma} = \begin{bmatrix} \bar{\Sigma}_{xx} & 0 \\ 0 & I \end{bmatrix}$ and $[0, I]K_{AM} = 0$ and $[I_{n_x}, 0]K_{AM} = K$, where $K = (A\bar{\Sigma}_{xx}C' + B\tilde{D}'_\epsilon)V_y^{-1}$ where $V_y = (C\bar{\Sigma}_{xx}C' + \tilde{D}_\epsilon\tilde{D}_\epsilon + \tilde{D}_v\tilde{D}_v)$ and $\bar{\Sigma}_{xx} = A\bar{\Sigma}_{xx}A' + BB'$. □

Proof of Theorem **??**:

Proof. Let $\xi_t = [x'_t, \hat{x}'_{t|t}]'$. The H-period ahead covariance is given by $\text{Var}[y_{t+H}|y_t] = \sum_{h=0}^{H-1} \tilde{A}^h\tilde{B}\tilde{B}'(\tilde{A}^h)'$. Since $\tilde{A}$ has a single unit root, inherited from x_t, the variance is dominated by $\sum_{h=1}^{H-1} V_{\tilde{A}}e_1e'_1V_{\tilde{A}}^{-1}\tilde{B}\tilde{B}'V_{\tilde{A}}e_1e'_1V_{\tilde{A}}^{-1} = (H-1)V_{\tilde{A}}e_1e'_1V_{\tilde{A}}^{-1}\tilde{B}\tilde{B}'V_{\tilde{A}}e_1e'_1V_{\tilde{A}}^{-1}$. When x_t is a scalar, then $V_{\tilde{A}} = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1 \end{bmatrix}$ and $V_{\tilde{A}}^{-1} = \begin{bmatrix} \sqrt{2} & 0 \\ -1 & 1 \end{bmatrix}$ so that the variance behaves approximately as $\text{Var}[y_{t+H}|y_t] \approx (H-1)\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\tilde{B}\tilde{B}'\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. $\text{Cov}[x_{t+H}, \hat{x}_{t+H|t+H}|y_t] \approx (H-1)[0,1]\,\text{Var}[y_{t+H}|y_t][1,0]' = (H-1)[1,0]\tilde{B}\tilde{B}'[1,0]' = (H-1)[\tilde{B}\tilde{B}']_{1,1}$. This covariance equals also the variance of the state: $\text{Var}[x_{t+H}|y_t] \approx (H-1)[1,0]\,\text{Var}[y_{t+H}|y_t][1,0]' = (H-1)[1,0]\tilde{B}\tilde{B}'[1,0]' \approx \text{Cov}[x_{t+H}, \hat{x}_{t+H|t+H}|y_t]$. Similarly, $\text{Var}[\hat{x}_{t+H|t+H}|y_t] \approx (H-1)[0,1]\,\text{Var}[y_{t+H}|y_t][0,1]' = (H-1)[1,0]\tilde{B}\tilde{B}'[1,0]' \approx \text{Cov}[x_{t+H}, \hat{x}_{t+H|t+H}|y_t]$. Thus, the correlation thus asymptotes to one as $H \to \infty$. □

B Holston, Laubach and Williams (2023) estimation details

Table 4: Prior and Posterior Distributions

Parameter	Prior Distribution	Prior Mean	Prior Std	Posterior Mean	Posterior Std	Posterior 90% HPD
a1	Normal	1.5000	0.5000	1.4936	0.0935	[1.3447, 1.6487]
rho	Beta	0.9000	0.1000	0.9302	0.0232	[0.8936, 0.9690]
a_r	Normal	-0.2000	0.2000	-0.0542	0.0210	[-0.0879, -0.0201]
bpi	Beta	0.6000	0.2000	0.6824	0.0595	[0.5841, 0.7799]
c	Normal	1.0000	0.5000	1.2204	0.4096	[0.6129, 1.9458]
by	Normal	0.1000	0.5000	0.0800	0.0361	[0.0197, 0.1355]
kappa2020	Inv. Gamma	5.0000	3.0000	4.5543	1.7125	[2.0606, 7.2668]
kappa2021	Inv. Gamma	2.0000	3.0000	1.8985	1.3375	[0.5345, 3.5124]
kappa2022	Inv. Gamma	2.0000	3.0000	1.8731	1.3340	[0.5282, 3.3723]
phi	Normal	0.0000	1.0000	-0.1230	0.0296	[-0.1716, -0.0745]
se_ygap	Inv. Gamma	0.5000	0.4000	0.4507	0.0641	[0.3448, 0.5564]
se_pi	Inv. Gamma	0.5000	0.4000	0.7850	0.0349	[0.7273, 0.8417]
se_ystar	Inv. Gamma	0.5000	0.4000	0.3597	0.0917	[0.2077, 0.5053]
se_gg	Inv. Gamma	0.1000	0.4000	0.1468	0.0681	[0.0463, 0.2451]
se_zz	Inv. Gamma	0.1000	0.4000	0.0916	0.0915	[0.0239, 0.1595]
se_rr	Inv. Gamma	0.5000	2.0000	1.1897	0.0519	[1.1041, 1.2735]
se_strobs	Inv. Gamma	0.5000	2.0000	3.8212	0.1666	[3.5455, 4.0900]
se_yobs	Inv. Gamma	0.5000	2.0000	0.2267	0.0442	[0.1532, 0.2987]

C UC3 estimation details

Table 5: Prior and Posterior Distributions

Parameter	Prior Distribution	Prior Mean	Prior Std	Posterior Mean	Posterior Std	Posterior 90% HPD
a_1	Normal	1.5000	0.5000	1.6212	0.0621	[1.5192, 1.7235]
ρ	Beta	0.5000	0.1500	0.9557	0.0098	[0.9401, 0.9726]
b_1	Normal	0.1000	0.5000	0.1274	0.1016	[0.0002, 0.2635]
b_2	Normal	0.0500	0.5000	0.0934	0.1200	[-0.1063, 0.2915]
c_1	Normal	-0.5000	0.5000	-0.4878	0.1061	[-0.6478, -0.3158]
c_2	Normal	-0.2500	0.5000	-0.2105	0.0938	[-0.3382, -0.0563]
σ_c	Inv. Gamma	0.5000	0.4000	0.3674	0.0549	[0.2745, 0.4562]
σ_{y*}	Inv. Gamma	0.5000	0.4000	0.9677	0.0467	[0.8964, 1.0498]
σ_{u*}	Inv. Gamma	0.1000	0.4000	0.0549	0.0183	[0.0255, 0.0824]
$\sigma_{\pi*}$	Inv. Gamma	0.5000	0.4000	0.5901	0.0606	[0.4972, 0.6946]
σ_y	Inv. Gamma	0.5000	2.0000	0.1487	0.0291	[0.1044, 0.1956]
σ_u	Inv. Gamma	0.2000	2.0000	0.0668	0.0105	[0.0494, 0.0839]
σ_π	Inv. Gamma	0.2500	1.0000	0.4532	0.0542	[0.3617, 0.5386]

D Prior-posterior comparison

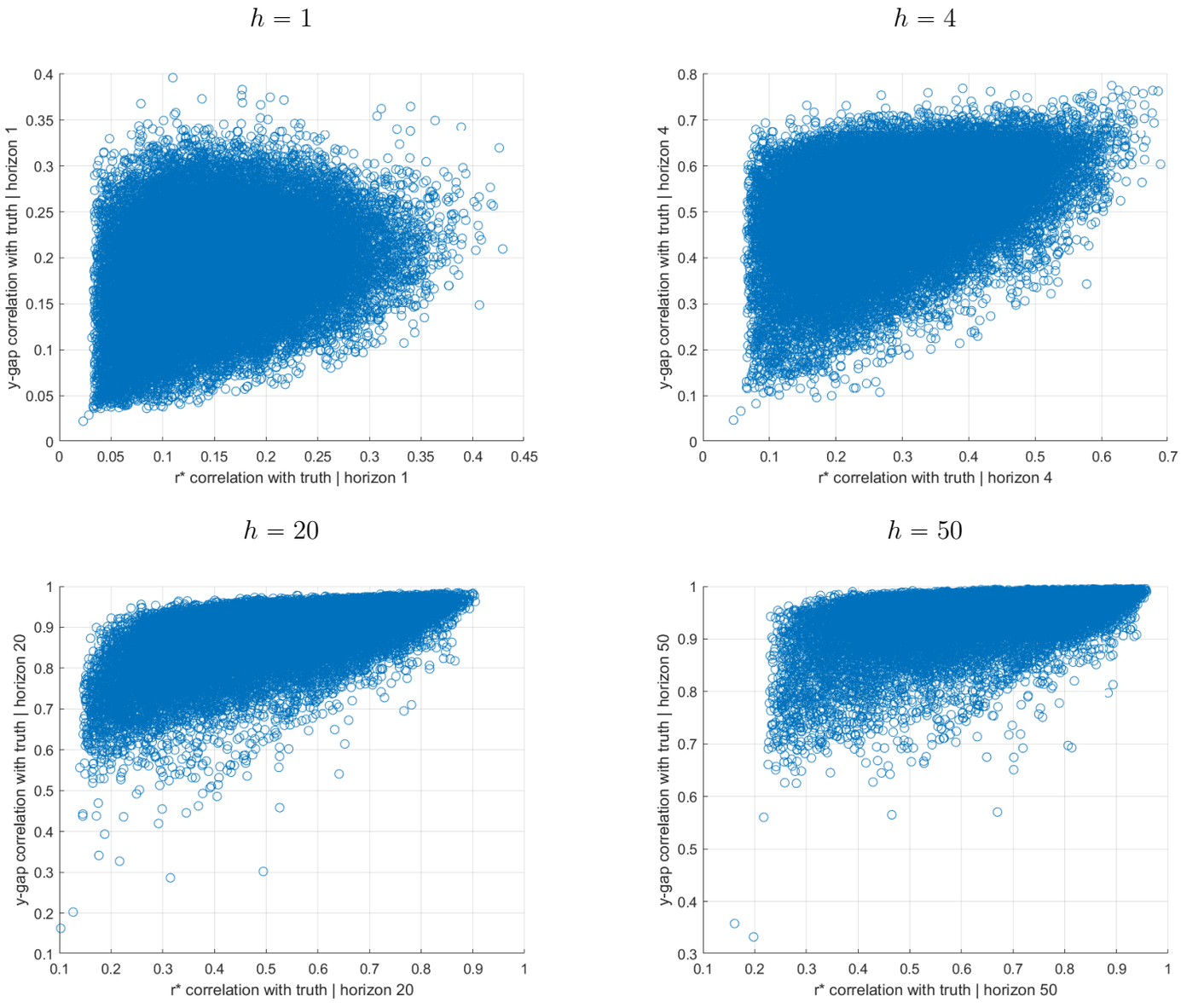


Figure 20: Correlation of $\hat{r}_t^*$ with the truth at various horizons vs correlation of the output gap estimate with the truth (posterior)

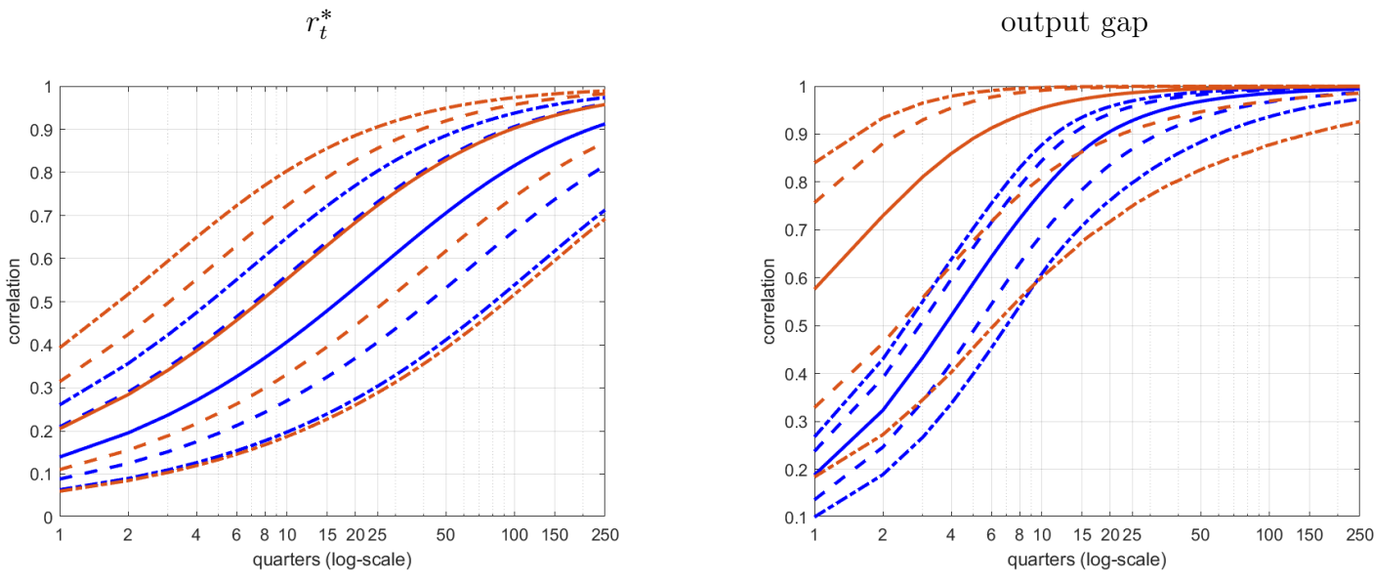


Figure 21: Prior vs posterior correlation of state estimate with the truth as a function of the horizons: $\hat{r}_t^*$ and output gap (prior vs posterior)

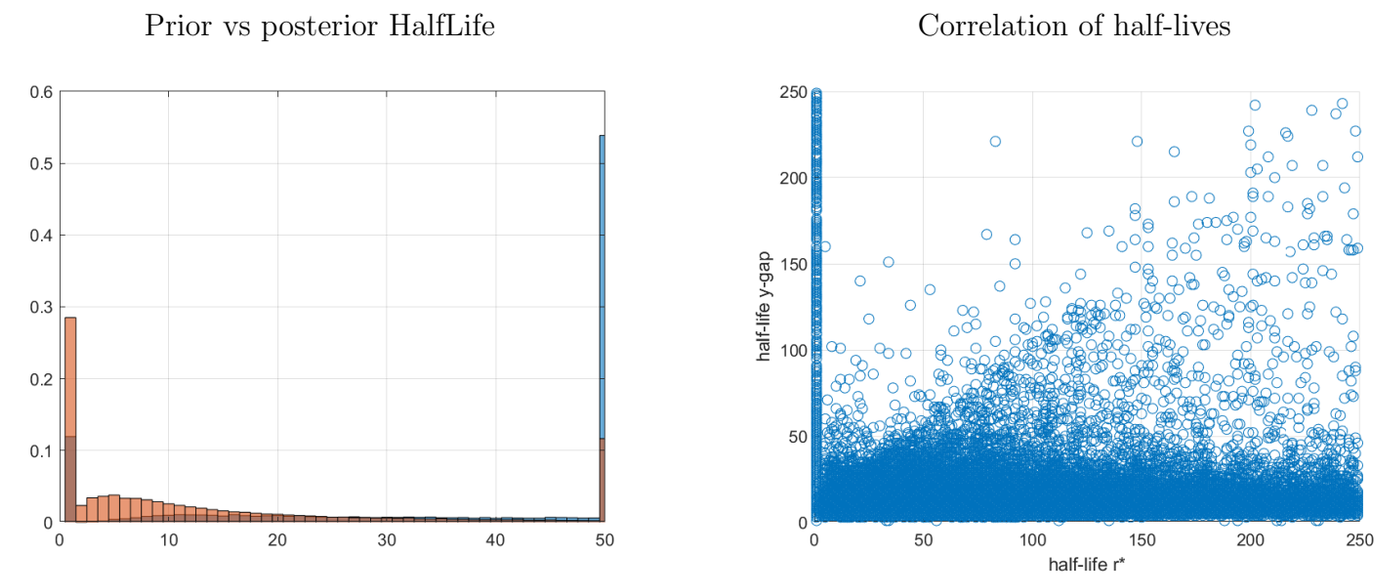


Figure 22: Half-life of $\hat{r}_t^*$ errors and correlation of half-life of r_t^* errors with the half-life of the output gap errors (prior vs posterior)

Parameter	output growth	inflation
sum of lag coeffs	NaN	NaN
1st lag coeffs	NaN	NaN
r-y	NaN	NaN
lagged inflation	NaN	NaN
slope PC	NaN	NaN
COVID-loading	NaN	NaN
SE y-gap	NaN	NaN
SE infl	NaN	NaN
SE potential y	NaN	NaN
SE z	NaN	NaN
SE growth	NaN	NaN
SE rr	NaN	NaN

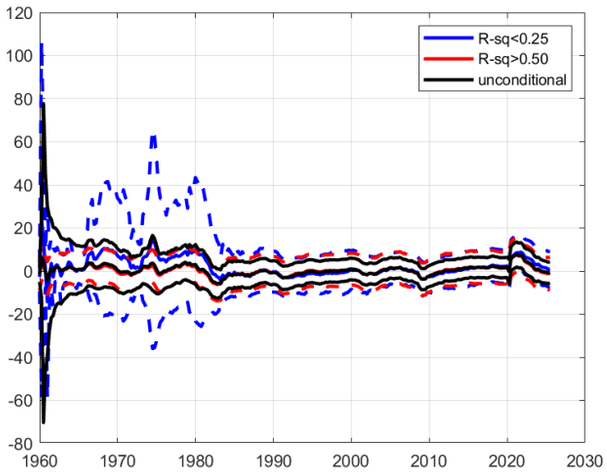
Table 6: Correlation of 1-quarter-ahead RMSEs with parameters

D.1 Bayes Factor

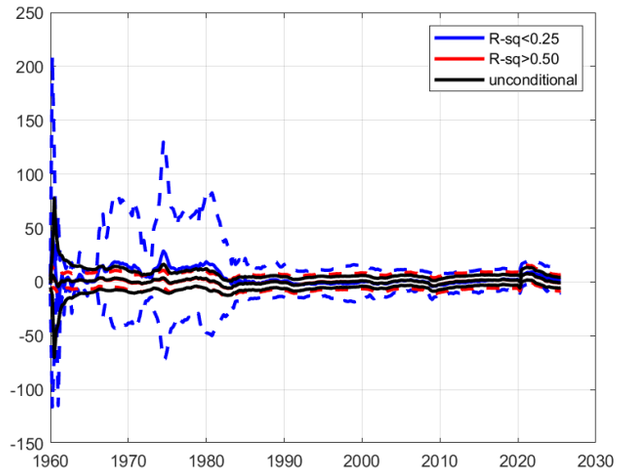
We compute the Bayes factor as the ratio of the posterior probability of a reliability exceeding some threshold relative to the prior probability of exceeding said threshold. For simplicity, we focus on the correlation at some horizon exceeding 0.5 (or 0.25).

output gap

Horizon 8

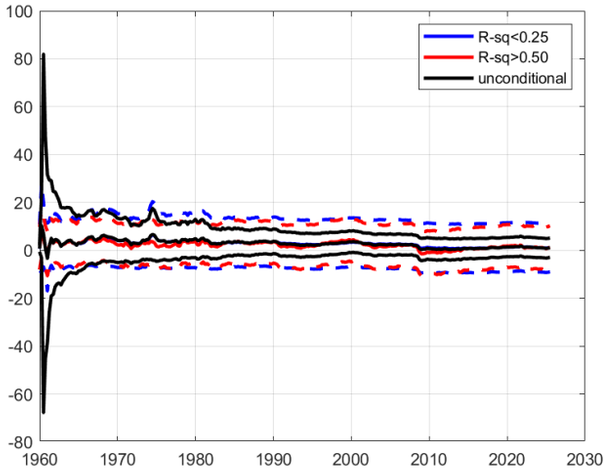


Horizon 12



r^*

Horizon 12



Horizon 16

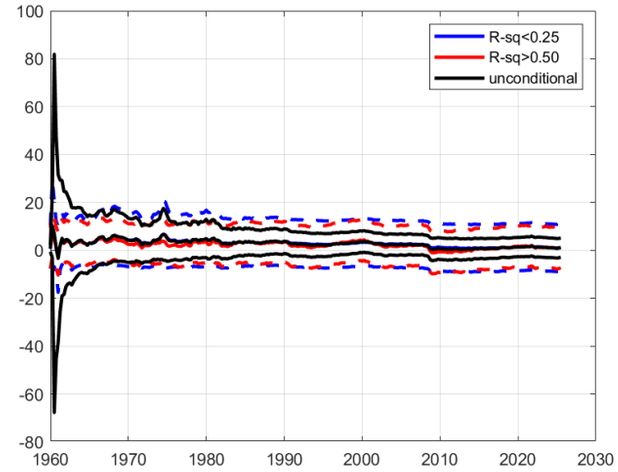


Figure 23: Tracking estimates for the output gap and r^* conditional on reliability at different horizons

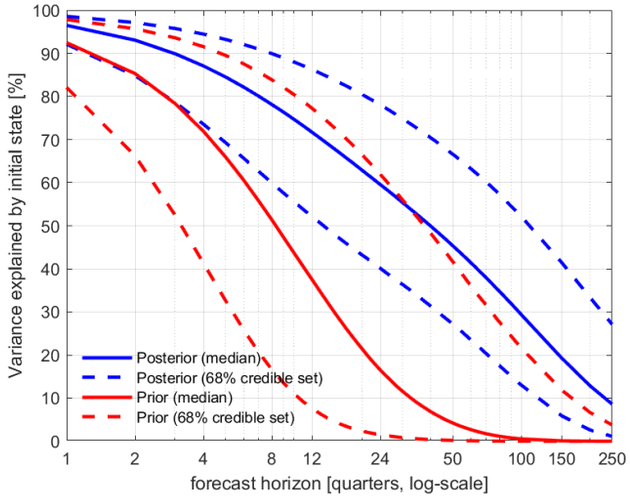
	Half-life			correlation				
Parameter	unconditional	half-life>1	Parameter	H=1	H=4	H=10	H=20	H
sum of lag coeffs	0.02	0.20	sum of lag coeffs	-0.33	-0.33	-0.32	-0.32	-(
1st lag coeffs	0.02	0.17	1st lag coeffs	-0.24	-0.24	-0.23	-0.23	-(
r-y	-0.07	0.42	r-y	-0.49	-0.49	-0.50	-0.50	-(
lagged inflation	-0.03	0.11	lagged inflation	-0.11	-0.11	-0.11	-0.12	-(
slope PC	0.06	-0.25	slope PC	0.27	0.27	0.28	0.28	(
COVID-loading	0.00	0.02	COVID-loading	-0.02	-0.02	-0.02	-0.02	-(
SE y-gap	-0.02	0.08	SE y-gap	-0.10	-0.10	-0.10	-0.10	-(
SE infl	-0.02	0.07	SE infl	-0.07	-0.07	-0.07	-0.07	-(
SE potential y	0.04	0.06	SE potential y	-0.13	-0.13	-0.13	-0.12	-(
SE z	0.08	-0.07	SE z	-0.03	-0.02	-0.01	-0.00	(
SE growth	-0.12	-0.49	SE growth	0.71	0.71	0.69	0.67	(
SE rr	-0.00	-0.01	SE rr	0.01	0.01	0.01	0.01	(

Table 7: Correlation of half-life and correlation with parameters: r_t^*

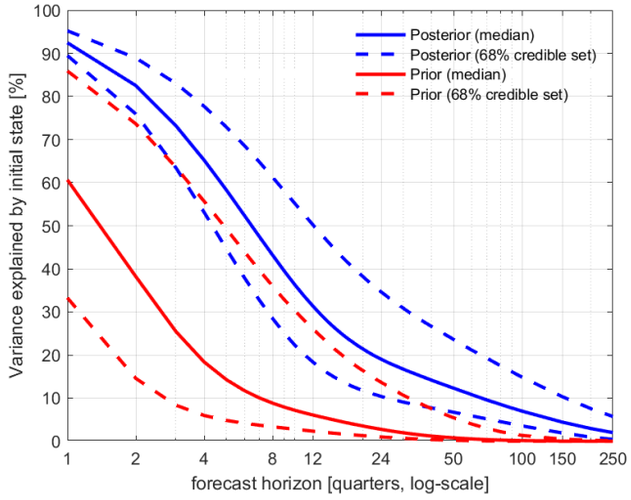
Parameter	unconditional	half-life>1	Parameter	H=1	H=4	H=10	H=20	H
sum of lag coeffs	0.49	0.53	sum of lag coeffs	-0.33	-0.52	-0.43	-0.19	(
1st lag coeffs	-0.19	-0.21	1st lag coeffs	-0.04	0.09	0.07	-0.06	-(
r-y	-0.29	-0.33	r-y	0.06	-0.03	-0.43	-0.64	-(
lagged inflation	0.26	0.29	lagged inflation	-0.35	-0.32	-0.23	-0.12	-(
slope PC	-0.72	-0.81	slope PC	0.86	0.76	0.51	0.23	(
COVID-loading	0.10	0.11	COVID-loading	-0.02	-0.07	-0.08	-0.04	(
SE y-gap	0.33	0.36	SE y-gap	-0.02	-0.09	-0.21	-0.13	-(
SE infl	0.18	0.20	SE infl	-0.23	-0.20	-0.14	-0.08	-(
SE potential y	-0.30	-0.32	SE potential y	0.09	0.00	0.01	-0.03	-(
SE z	0.21	0.22	SE z	0.05	-0.14	-0.20	-0.20	-(
SE growth	-0.21	-0.21	SE growth	-0.03	0.10	0.24	0.21	(
SE rr	-0.01	-0.01	SE rr	0.01	0.02	0.07	0.08	(

Table 8: Correlation of half-life and correlation with parameters: output gap

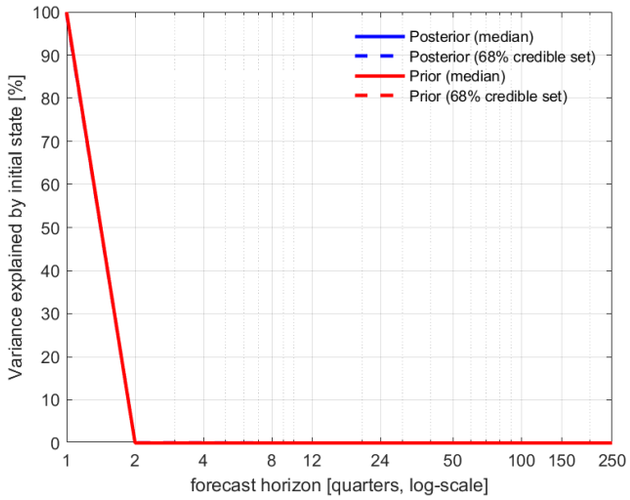
r_t^*



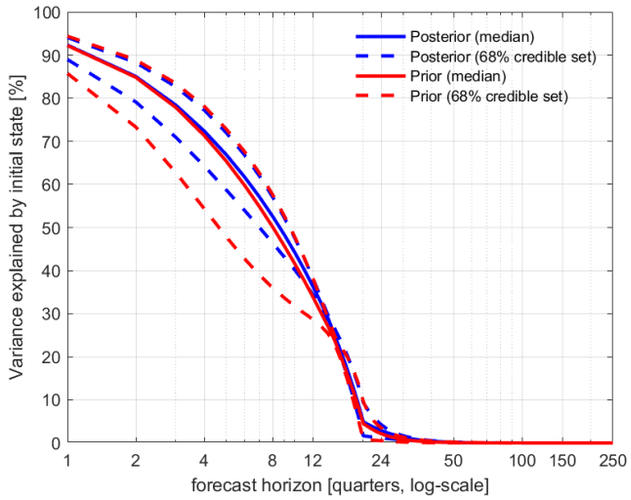
output gap



1-quarter change in r_{t*}



5-year change in r_t^*



Plotted is $100 - 100 \times \frac{\left[\sum_{h=1}^{H}(A-KC)^{h-1}(B-KD)(B-KD)'((A-KC)')^{h-1}\right]_{ii}}{[\Omega_{t|t}]_{ii}}$ against H.

Figure 24: Contribution of the initial to the state estimation uncertainty by horizon and state of interest

E Assessing models: Application to UC3

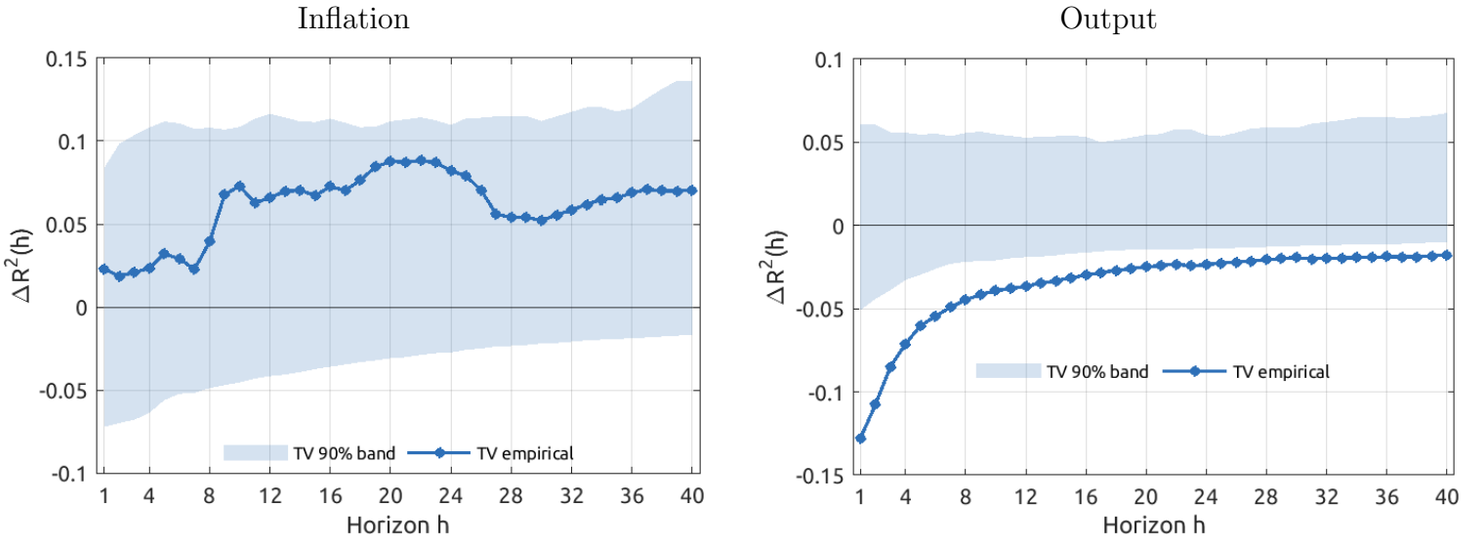


Figure 25: Gap between the theoretical and empirical R^2 in (6.1) for inflation and output at various horizons: Posterior mean vs simulated distribution.

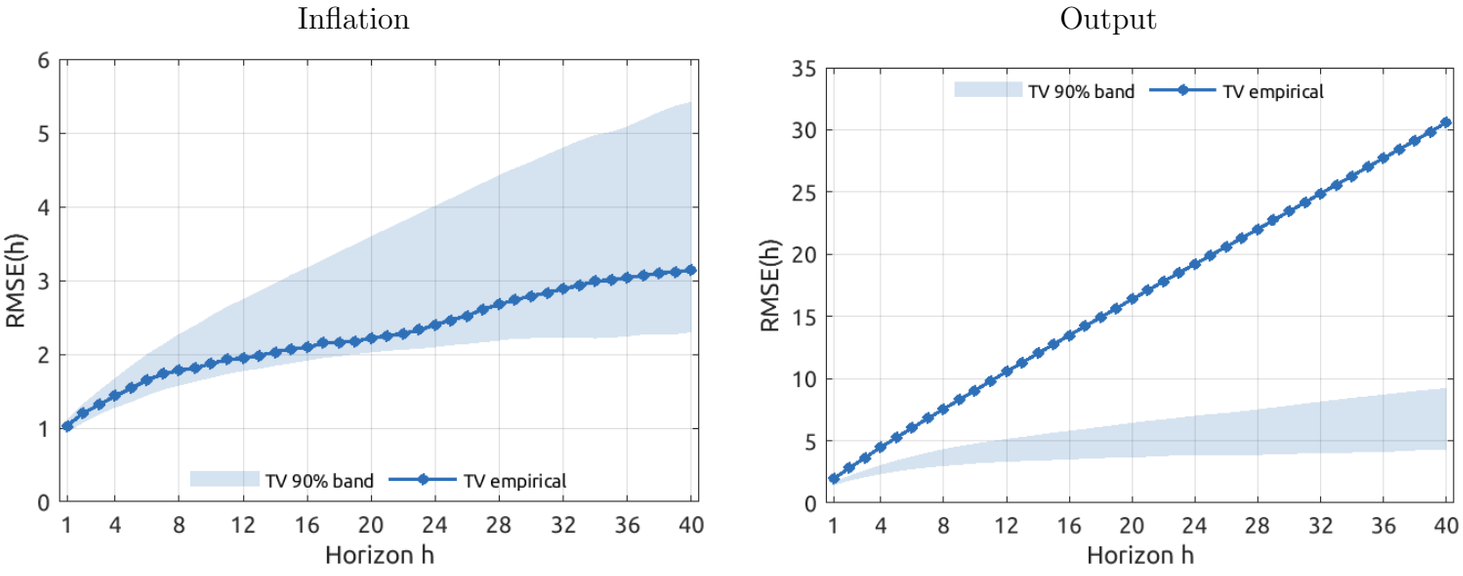


Figure 26: RMSE for inflation and output growth at various horizons: Posterior mean vs simulated distribution.