Nonlinear Expectations: Making Sense of Professional Forecasts

Julio Ortiz *

First draft: September 2019

Latest draft: May 2020

View Latest Version Here

Abstract

This paper introduces a novel finding about professional forecasters: the same respondent simultaneously over- and underreacts to distinct macroeconomic variables. Whereas standard models struggle to make sense of this fact, I show that simultaneous over- and underreactions can arise in nonlinear environments. In such contexts, rational forecasters are unable to obtain an optimal and exact estimate of the state via the Kalman filter. I consider stochastic volatility as the source of nonlinearities and assume that forecasters must approximate the state by selecting an approximation function from a set of costly alternatives. According to my model, over- and underreactions depend crucially on the the underlying signal-to-noise ratio. In particular, the scope for overreactions is decreasing in the signal precision. I conclude that error predictability is not prima-facie evidence against rationality.

Keywords: Rational expectations. Noisy information. Forecasting. Stochastic volatility. Nonlinear filtering.

JEL Codes: C11, C13, C15 D83, D84

*Boston University, Department of Economics, 270 Bay State Road, Boston, MA 02215; Phone: 201-230-1960; Email: jlortiz@bu.edu.

1 Introduction

Professional forecasts exhibit error predictability. While several theories have been devised to explain these apparent deviations from rationality, the literature on expectation formation is unable to successfully make sense of simultaneous over- and underreactions to different macroeconomic variables. I argue that the key to matching this empirical fact lies in relaxing a restriction on how the latent state is allowed to evolve. Macroeconomic and financial time series have been found to exhibit complex dynamics such as stochastic volatility, structural breaks, and regime switching. At present, most models of expectations formation assume a linear environment. This provides the benefit of tractability, however, when taking the theoretical implications to the data, these models attribute any deviation from constant gain Kalman filtering to non-rational behavior.

Addressing this misspecification can make error predictability compatible with rationality. This is because nonlinear models no longer imply a simple analytical solution to the minimum mean square error estimate. Instead, to generate predictions forecasters must first approximate the nonlinear state. If there are costs to implementing approximation methods, then the choice of which approximation to use is not immediately obvious. Subject to such costs, forecasters can optimally adopt different approximation functions. As a result, revisions can hold predictive power over errors, and conventional efficiency tests would not necessarily detect non-rational behavior.

Survey data is predominantly used to test theories of expectation formation. In this paper, I make use of the Survey of Professional Forecasters (SPF) which provides a panel of multi-horizon forecasts across several macroeconomic variables. The literature has documented evidence of error predictability at both the forecaster- (individual) level as well as at the consensus- (aggregate) level across several surveys.[1] At the consensus-level, underreactions tend to prevail. Interestingly, at the forecaster-level there is evidence of coincident over- and underreactions among forecasters. I show how prevalent this phenomenon is by computing the share of forecasters that simultaneously overreact to one variable and underreact to another in the SPF.

Theories such as those of forecast smoothing and strategic complementarities have been offered to explain individual and aggregate underreaction. At the same time, models of diagnostic expectations and overconfidence can deliver individual overreaction and aggregate underreaction. A unified theory that accommodates

[1] Examples other than the SPF include the Livingston survey, the Michigan Survey of Consumers, the NY Fed Survey of Consumer Expectations, Blue Chip forecasts, the ECB Survey of Professional Forecasters, and the daily Focus Survey from the Central Bank of Brazil, among others.

either behavior at the individual-level, however, is missing. Simultaneous over- and underreaction prompts several fundamental questions about expectations formation. Are professional forecasters (presumably the most informed private agents in the economy) rational? Does a behavioral bias govern the manner in which expectations are formed? As policymakers increasingly pursue expectations-based policies such as forward guidance, taking a step toward reconciling theories of expectations formation with the data is of first-order importance.

Against this backdrop, I present a nonlinear noisy information model. Specifically, stochastic volatility coupled with heterogeneity in use of approximation functions can replicate important features of the data. Rather than obtaining an exact solution to the optimal inference problem, forecasters must approximate the posterior distribution. They may choose from a finite set of approximation methods. The available methods vary in their level of sophistication, and adopting a given approximation function is subject to a random cost. These costs are increasing with the sophistication of the approximation. Forecasters generate a prediction that minimizes the sum of their mean squared errors and adoption costs.

I take a systematic approach to building this model by first generalizing a standard linear noisy information rational expectations model. This is done in order to incorporate a number of extensions such as a multivariate state as well as public and private signals. I focus on forecasts made about public signals because while forecasters formulate expectations about unobserved fundamentals, *they report predictions about public signals observed with a lag*. For instance, forecasters observe past inflation and formulate expectations conditional on their respective information sets. These inflation expectations are ultimately what forecasters report when surveyed. Assuming that inflation is a function of some unobserved state, then this variable is best modeled as a public signal. The flexible framework presented implies that one can no longer extract an estimate of information rigidity by merely projecting ex-post errors on ex-ante revisions. Furthermore, I verify that linearity and rationality jointly imply error orthogonality regardless of the generalizations made to state dynamics or signal structure.

After this, I modify the general model by incorporating time varying volatility into the state transition equation. I show that deviating from the simple linear state space model can deliver error predictability among otherwise rational forecasters who face costs to adopting approximation functions. Such a model implies that current period forecasts can be expressed as a sum of the optimal minimum mean square error forecast and an approximation error. To extract additional intuition for the sources of over- and underreactions, I consider a stylized version of the model. For simplicity, I assume that forecasters can either approximate the state via a naive linear approximation that uses the Kalman filter, or via an asymptotically efficient approximation method (the particle filter). The latter approximation method is more costly than the former.[2] There are several motivations for modeling costs that increase with level of sophistication such as the cognitive cost to generating a forecast that uses a more complex approximation function. Alternatively, a more sophisticated approximation method may require additional computing time.

This stylized model is able to speak to the relation between errors and revisions at either level of aggregation. I find that the covariance between errors and revisions depends importantly on the data generating process. Put another way, there are certain features inherent to a given macroeconomic time series that explain why forecasters appear to either over- or underreact to that particular variable. Simulation results suggest that overreactions arise when the signal-to-noise ratio is low. On the other hand, when there is less noise characterizing the system, the model delivers underreactions.

I verify these implications in the data by exploiting the cross-section of macroeconomic variables for which forecasters report predictions in the SPF. In addition to this, I calibrate the stylized model for real GDP and the unemployment rate and find that the implied signal-to-noise ratio is consistent with the simulation results. Taken together, my findings suggest that error predictability is not sufficient to reject rationality in all contexts, but rather, it is evidence against *linear* rational expectations.

1.1 Related Literature

In their seminal paper, Coibion and Gorodnichenko (2015), henceforth CG, make sense of forecast error inefficiency while preserving the assumption of rationality. They do so by incorporating imperfect information into a standard linear rational expectations model. Using consensus-level data, CG show that projecting *ex-post* forecast errors on *ex-ante* forecast revisions delivers an estimate of information rigidity.

More recently, many studies have used forecaster-level data to test for rationality.[3] In doing so, much of this literature preserves the linearity assumption made in CG, and ultimately rejects rational expectations even under imperfect information. Running the same regression as CG at the forecaster-level delivers non-zero point estimates for most macroeconomic variables. To make sense error predictability at the individual-level

[2] I have also experimented with a local linear approximation method such as the Extended Kalman filter, and the Unscented Kalman filter which makes use of specially chosen mass points. Though I do not provide an analytical result here, one could in princple expand the set of approximation functions arbitrarily by indexing an approximation function by the number of particles specified in a particle filter. I provide brief summaries on these methods in Appendix C.

[3] Examples include Bordalo et al. (2019), Fuhrer (2018), Dovern et al. (2015), Andrade and Le Bihan (2013), Broer and Kohlhas (2018), Burgi (2018).

while also matching the CG finding of underreactions at the aggregate level, several theories of non-rational expectations have been proposed. [4]

Moreover, the model in this paper is in the spirit of Branch (2004), Evans and Ramey (1992) and Brock and Hommes (1997) who define adaptively rational equilibrium dynamics (ARED). Branch (2004) was the first to introduce this concept to expectations formation. This paper adds to those insights in key ways. First, I present more complex dynamics for the state variable. Introducing nonlinearities, such as stochastic volatility, provides a stronger justification for the use of different predictor (approximation) functions. Second, I allow for heterogenous expectations whereas Branch (2004) assumes homogeneous predictions among all who adopt a specific predictor function.[5] Taken together, my model is able to reproduce the empirical facts that are present in survey data both at the individual and aggregate levels.

While a discussion of nonlinearities has generally been absent in the survey expectations literature, the finance literature has previously tied nonlinear dynamics to error predictability. For instance, Lewis (1989) considers error predictability concerning dollar forecasts in the context of a structural break. Veronesi (1999) finds that over- and underreactions arise in a regime switching model of asset pricing. More recently, Lansing et al. (2019) attribute the predictability of excess returns to either volatility or deviations from rationality. To this end, my paper also relates to the literature on volatility in macroeconomics. See for instance, Justiniano and Primiceri (2007), Kim and Nelson (1998), McConnell and Perez-Quiroz (2000), and Stock and Watson (2002). Finally, due to the assumptions imposed on the state dynamics, this paper relates to the literature on nonlinear filtering. Several approximation methods have been devised in order to deal with nonlinearities in the evolution of a state variable. These methods include generalizations to Kalman filtering as well as importance sampling algorithms, among others.[6] A strand of this literature has recently formalized some basic efficiency properties of particle filtering.[7]

1.2 Roadmap

This paper is structured as follows. Section 2 presents previously documented facts about error predictability at the forecaster and consensus levels, as well as a novel fact on simultaneous over- and underreactions.

[4] For instance Bordalo et al. (2019) rule out rationality in favor of diagnostic expectations. Other studies such as Fuster et al. (2010) argue in favor of models featuring misperception at long horizons. Daniel et al. (2003) argues for a model of overconfidence while Broer and Kohlhas (2018) present a model of relative overconfidence.

[5] Heterogeneity in that model comes from idiosyncratic "trembles" in the reported prediction.

[6] Julier and Uhlman (2004) develop a Kalman filter for nonlinear settings while Doucet and Johansen (2009) discuss particle filtering methods.

[7] See Crisan and Doucet (2002) and Hu et al. (2011)

Section 3, presents a generalization of the standard model of expectations formation in the literature. Here I show that linearity is at the heart of error orthogonality. Section 4 incorporates nonlinear dynamics into the generalized model. Section 5 provides simulation results on a stylized version of this model. Section 6 reports empirical evidence in favor of a nonlinear model. Section 7 parameterizes the stylized model, and Section 8 concludes.

2 Evidence from Survey Data

The SPF is a quarterly survey provided by the Federal Reserve Bank of Philadelphia. The survey began in 1968Q4 and provides forecasts from several forecasters across a number of macroeconomic variables including real GDP, inflation, and unemployment, over many horizons, h. The variables of interest in this paper are the forecast error and the forecast revision. To construct forecast errors from forecaster i about variable x, $FE^i_{t+h,t} = x_{t+h} - x^i_{t+h|t}$, I take the difference between the realized real-time value for x at $t+h$ and the forecaster's h-step ahead prediction generated at time t. This convention is in keeping with the literature as using final estimates reported by the Bureau of Economic Analysis several quarters into the future could reflect data revisions. To compute forecast revisions, I exploit the term structure of forecasts generated by the survey respondents. Specifically, $FR^i_{t,t-1} = x^i_{t+h|t} - x^i_{t+h|t-1}$. This requires making use of the $h-$step ahead forecasts formulated in periods t and $t-1$. In other words, I consider the fixed horizon, h and take the difference between two adjacent forecasts.

2.1 Error Predictability at the Forecaster and Consensus Levels

CG present the following testable implication at the consensus-level which holds for an arbitrary horizon, h

$$FE_{t+h,t} = \alpha_0 + \alpha_1 FR_{t+h,t} + \epsilon_t \tag{1}$$

CG find that in the data, $\alpha_1 > 0$ for most variables which indicates that consensus forecasts underreact to new information. The gradual adjustment of expectations that arises from the optimal filtering of noisy signals leads to the inertia in expectations formation.

More recently, Bordalo et al. (2019) estimate the same regression at the forecaster-level:

$$FE^i_{t+h,t} = \beta_0 + \beta_1 FR^i_{t+h,t} + \varepsilon^i_t \tag{2}$$

	Nowcast		One-Quarter Ahead		Two-Quarters Ahead	
	β_1	α_1	β_1	α_1	β_1	α_1
Estimate	-0.317***	0.569***	-0.231**	1.011***	-0.344***	0.565**
	(0.050)	(0.128)	(0.067)	(0.201)	(0.058)	(0.272)
Observations	65070	2323	54067	2309	52220	2295

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 1: Pooled OLS Forecast Error Predictability Regressions

Note: The table reports the estimated coefficients of forecast error predictability at the current, one-, and two-quarter ahead horizons. Across all horizons, column (1), refers to the forecaster-level errors-on-revisions regression. Column (2) refers to the consensus-level errors-on-revisions regression. Standard errors for forecaster-level regressions are as in Driscoll and Kraay (1998), while Newey-West standard errors are used for aggregate-level specifications. Data used for estimation come from SPF.

They find that $\beta_1 < 0$ for most macroeconomic series. The interpretation is that forecasters overreact to new information. Bordalo et al. (2019) explain these empirical findings with a model of diagnostic expectations wherein agents place more weight on recent developments as they believe these to be more representative of the objective probability distribution.

2.2 Facts About Professional Forecasters

Error Predictability Regressions

Table 1 reports estimates of β_1 and α_1, using data from the SPF. The estimates are obtained via OLS regressions, pooling across both forecasters and macroeconomic variables. Estimates are reported for three different horizons. Across all horizons considered, it is clear that overreactions dominate at the individual-level, while underreactions arise at the aggregate-level.

In Table 2, I focus on the nowcast ($h = 0$), and estimate each testable implication variable-by-variable. The results point to individual overreactions for most variables (consistent with Bordalo et al (2019)), although error orthogonality cannot be rejected for some variables. Furthermore, unemployment exhibits underreactions at the individual-level. At the consensus-level, underreactions dominate, however, real consumption and real government spending (both federal and state/local) feature no error predictability.

Simultaneous Over- and Underreactions

Whereas these empirical facts can arise due to non-rational biases, existing theories struggle to explain why a forecaster might simultaneously overreact to one variable and underreact to another. To get a sense of how common such an occurrence is, I estimate regression (2) for each forecaster i forecasting a specific

Variable	Mnemonic	β_1	α_1
Consumer price index	CPI	-0.085	0.868***
Employment	EMP	-0.123	0.564***
Housing starts	HOUSING	0.063	0.359***
Industrial production	IP	-0.147*	0.513***
Nominal GDP	NGDP	-0.310***	0.421**
GDP Deflator	PGDP	-0.363***	0.350**
Real consumption	RCONSUM	-0.401***	0.098
Real federal government spending	RFEDGOV	-0.483***	0.377
Real GDP	RGDP	-0.264***	0.350**
Real nonresidential investment	RNRESIN	-0.499**	0.362
Real residential investment	RRESINV	-0.234***	0.925***
Real state/local government spending	RSLGOV	-0.660***	-0.381
3-month Treasury bill	TBILL	0.010	0.178***
10-year Treasury bond	TBOND	0.020	0.154***
Unemployment rate	UNEMP	0.082**	0.247***

* p<0.10, ** p<0.05, *** p<0.01

Table 2: Pooled OLS Regressions at $h = 0$, by Variable

Note: The table reports the OLS coefficients from errors-on-revisions regressions across 15 macroeconomic variables reported in the Survey of Professional Forecaters. Column (3) reports the coefficient in front of the revision at the forecaster-level while column (4) reports the analogous coefficient using consensus-level data. The errors and revisions are for current period forecasts ($h = 0$).

variable j. This delivers an $N_i \times N_j$ matrix of estimates $\widehat{\beta}_{1,ij}$. I keep only those estimates that are significant at 5% level. I then fix a pair of SPF variables j and k, and compute the number of forecasters such that $\widehat{\beta}_{1,ij} < 0$ and $\widehat{\beta}_{1,ik} > 0$. Finally, I count the number of forecasters for whom this condition holds and normalize by the number of total forecasters reporting predictions about variables j and k. More formally, I estimate a matrix P whose elements are p_{jk} with

$$p_{jk} = \frac{\sum_i \mathbb{1}\left(\widehat{\beta}_{1,ij} < 0 \text{ and } \widehat{\beta}_{1,ik} > 0\right)}{\min\{N_j, N_k\}}$$

where N_x denotes the number of forecasters providing predictions of variable x and $\mathbb{1}(\cdot)$ is the indicator function. The elements of matrix P, therefore, tell us the share of forecasters who simultaneously overreact to the row variable and underreact to the column variable. When p_{jk} is close to one, this means that nearly all forecasters overreact to variable j and underreact to variable k. On the other hand, when p_{jk} is close to zero, then almost no forecaster overreacts to variable j while also underreacting to variable k.

Figure 1 reports the results from this exercise. The figure verifies that a given forecaster tends to overreact

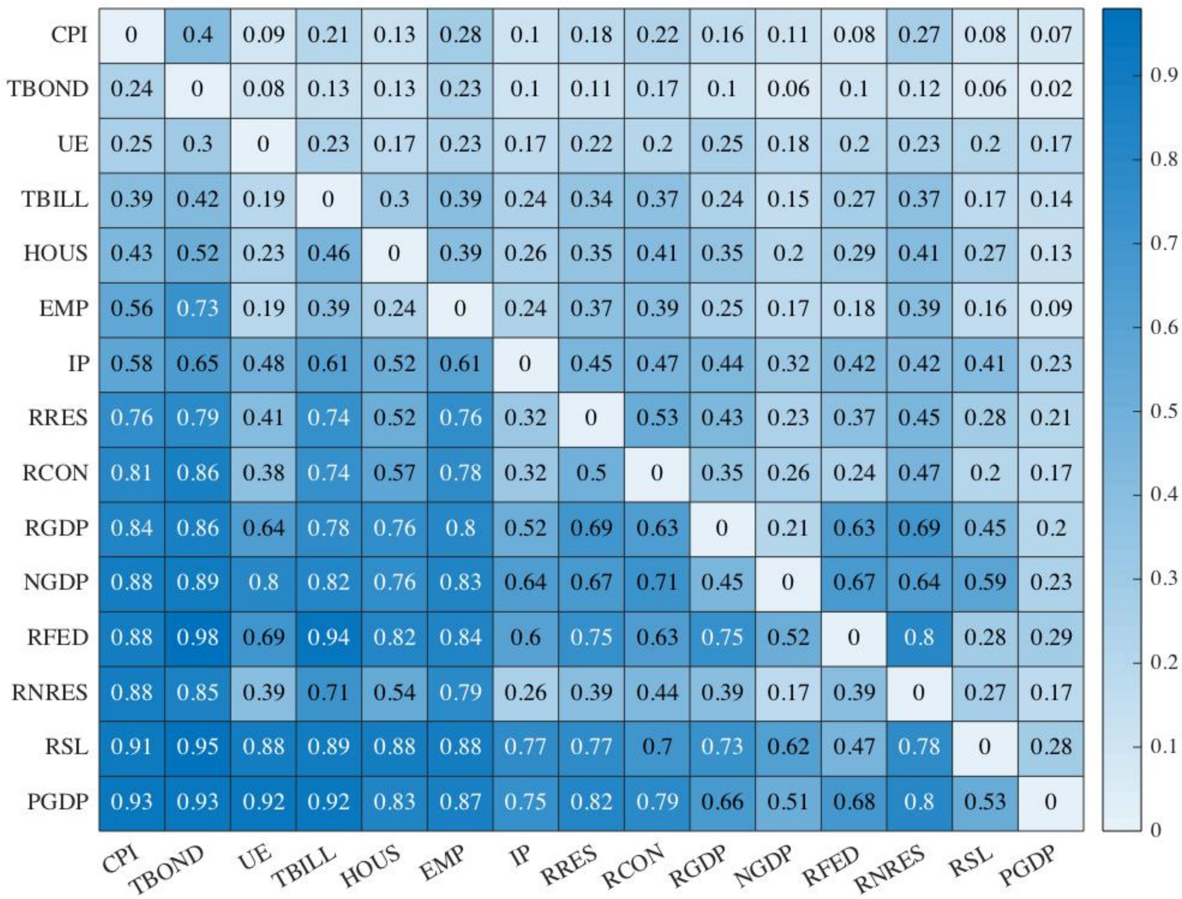


Figure 1: Frequency of Over- and Underreaction by the Same Forecaster

Note: This figure displays the share of forecasters who overreact to the row variable and simultaneously underreact to the column variable.

to some variables and underreact to others. This is particularly pronounced when considering overreactions to real variables. For instance, 84% of professional forecasters exhibit overreactions when forecasting real gross domestic product (RGDP) while simultaneously underreacting to their predictions about consumer prices (CPI). Moreover, it is less common for forecasters to overreact to financial variables as evidenced by the lighter shades for the 3-month treasury bills (TBILL) and 10-year bonds (TBOND) rows. In order to understand how individuals formulate their predictions of the future, a theory of expectations formation must take into account that a single agent may overreact and underreact to different variables.

3 Generalized Linear Baseline Model

Before presenting the nonlinear model, I generalize the standard model in the literature and provide some analytical results about error predictability. Consider a linear Gaussian state space model. Suppose there are n latent state variables and m exogenous signals.

$$\mathbf{s_t} = \mathbf{A s_{t-1}} + \mathbf{B w_t} \tag{3}$$
$$\mathbf{z_t^i} = \mathbf{C s_t} + \mathbf{D v_t^i}$$

Note that $\mathbf{s_t}$ is an $n \times 1$ vector, $\mathbf{A}$ is $n \times n$, $\mathbf{B}$ is $n \times n$ and $\mathbf{w_t}$ is $n \times 1$. Furthermore, $\mathbf{z_t}$ is $m \times 1$, $\mathbf{C}$ is $m \times n$, $\mathbf{D}$ is $m \times m$ and $\mathbf{v_t^i}$ is $m \times 1$. There are no other restrictions placed on the model. In particular, $\mathbf{s_t}$ can be a vector of many different state variables, or lags of itself. $\mathbf{B}$ need not be a diagonal matrix. Furthermore, $\mathbf{z_t^i}$ can include an arbitrary finite number of observed signals. The noise vector $\mathbf{v_t^i}$ can include private or public noise.[8]

3.1 Testable Implications of General Linear Model

From the Kalman filter, the optimal state estimate is defined as

$$\mathbf{s^i_{t|t}} = \mathbf{s^i_{t|t-1}} + \kappa(\mathbf{z^i_t} - \mathbf{z^i_{t|t-1}}) \tag{4}$$

where κ is the (constant) Kalman gain. Since the state is unobservable, forecasters can only formulate predictions of the signals and assess the mistakes made with regard to these observables. The optimal forecast of the signal vector $\mathbf{z_t^i}$ is

$$\mathbf{z^i_{t+1|t}} = \mathbf{z^i_{t+1|t-1}} + \mathbf{CA}\kappa(\mathbf{z^i_t} - \mathbf{z^i_{t|t-1}}) \tag{5}$$

Forecast errors for the generalized linear model can be expressed as follows

$$\mathbf{z^i_{t+1}} - \mathbf{z^i_{t+1|t}} = (\mathbf{z^i_{t+1}} - \mathbf{z^i_{t+1|t-1}}) - \mathbf{CA}\kappa(\mathbf{z^i_t} - \mathbf{z^i_{t|t-1}}) \tag{6}$$

[8] I index this vector by i in general to allow for forecaster-specific signals.

Furthermore, the forecast revision is

$$\mathbf{z^i_{t+1|t}} - \mathbf{z^i_{t+1|t-1}} = \mathbf{CA}\kappa(\mathbf{z^i_t} - \mathbf{z^i_{t|t-1}}) \tag{7}$$

Using these expressions, one can derive the two testable implications presented in the previous section.

Proposition 1. *The generalized linear model implies the following:*

(i) $\beta_1 = 0$

(ii) $\alpha_1 = \mathbf{CA}(\mathbf{I} - \mathbf{C}\kappa)(\kappa\mathbf{C})^{-\mathbf{1}}(\mathbf{CA})^{-\mathbf{1}} > 0$

Proof. See Appendix A. □

The proofs are straightforward and detailed in Appendix A: (a) holds given the orthogonality condition that must be satisfied at the individual-level under rational expectations. Forecast error orthogonality implies that $\mathbb{E}\big[(\mathbf{z^i_t} - \mathbf{z^i_{t|t}})\mu\big] = \mathbf{0}$ for any μ residing in the forecaster's information set.[9] Put another way, rationality implies the optimal use of information so that no variable residing in one's information set may predict the forecast error. This very general model precludes the predictability of forecast errors at the individual-level. As a result, any such linear Gaussian model with mean square loss cannot generate error predictability, regardless of the signal structure.

Moreover, (b) is a generalization of the CG result. The extent to which the mean revision predicts mean errors is determined by the Kalman gain matrix and the matrix $\mathbf{C}$ which maps the underlying state to the observed signal vector. The generalized linear model nests the CG result. Letting $\mathbf{C} = 1$, $\mathbf{D} = \sigma_v$, $\mathbf{A} = \rho$ and $\mathbf{B} = \sigma_w$, it follows that $\alpha_1 = \frac{1-\kappa}{\kappa}$. In this limiting case, one can recover an estimate of information rigidity by projecting consensus errors on consensus revisions. Importantly, the signal structure must be such that $\mathbf{C} = 1$. If, instead, the elements of $\mathbf{C}$ include additional parameters, or there is common noise in the signal vector, then it is no longer possible to cleanly extract an the Kalman gain from a standard OLS regression.[10]

[9] Similarly, there is a revision orthogonality condition implied by rationality which states that $\mathbb{E}(\mathbf{z^i_{t|t}} - \mathbf{z^i_{t|t-1}})|\mathcal{I}^i_t) = 0$. See Pesaran and Weale (2006).

[10] See CG for a discussion of the bias in estimated information rigidities induced by public noise.

4 A Model of Heterogeneous Rational Approximations

A nonlinearity such as time-varying volatility complicates the problem as forecasters must now formulate expectations about the level and the volatility of the underlying state.

4.1 Nonlinear Filtering Problem

To this end, I do away with the linearity assumption made in Section 3 and instead suppose that the volatility of the state varies over time. This requires modifying the covariance matrix of innovations: $\mathbf{B} \to \mathbf{B_t}$. One can generally formulate the model as follows:

$$\begin{aligned}\mathbf{\bar{s}_t} &= F(\mathbf{\bar{s}_{t-1}}, \mathbf{\overline{w}_t}) \\ \mathbf{z_t^i} &= \mathbf{C\bar{s}_t} + \mathbf{Dv_t^i}\end{aligned} \tag{8}$$

The crucial difference between the linear model and this one is the expanded state space: $\mathbf{\bar{s}} = \begin{bmatrix}\mathbf{s_t} & diag(\mathbf{B_t})\end{bmatrix}^\top$. This implies that the state is no longer linear as the error now enters multiplicatively into the state. This nonlinearity is modeled by the function $F(\cdot)$ which governs the evolution of the state. While the state now exhibits stochastic volatility, the shocks remain normal, and the signal structure is unchanged. Hence, the measurement equation remains linear.[11] [12]

Whereas Kalman filtering delivers an exact optimal solution in a linear Gaussian environment, it is no longer optimal in this context. The reason for this is that the Kalman filter requires one to evaluate the expected value of $\mathbf{\bar{s}_t}$ conditional on the history of signals $\mathcal{Z}_t^i = \{\mathbf{z_1^i}, \ldots, \mathbf{z_t^i}\}$. This is made intractable due to the lack of knowledge about the underlying conditional distribution. To see this more clearly, consider the scalar case where the state is s_t and there is only a private signal available to the forecaster, z_t^i. The observation equation can be expressed as a conditional likelihood, $p(z_t^i|s_t)$ and the state evolution as $p(s_{t+1}|s_t)$. The optimal filter computes $p(s_t|\mathcal{Z}_t^i)$ from a predict-update procedure:

$$p(s_t|\mathcal{Z}_{t-1}^i) = \int p(s_t|s_{t-1})p(s_{t-1}|\mathcal{Z}_{t-1}^i)ds_{t-1} \quad \text{(Predict)}$$

$$p(s_t|\mathcal{Z}_t^i) = \frac{p(z_t^i|s_t)p(s_t|\mathcal{Z}_{t-1}^i)}{p(z_t^i|\mathcal{Z}_{t-1}^i)} \quad \text{(Update)}$$

[11] This could be generalized to a nonlinear measurement as well. I abstract away from this for simplicity.

[12] While I consider stochastic volatility, any nonlinearity can deliver the results presented in the paper. In particular, this model can also speak to unobserved changes in the persistence of macroeconomic time series.

where $p(z_t|Z^i_{t-1}) = \int p(z^i_t|s_t)p(s_t|Z^i_{t-1})ds_t$.

In a linear Gaussian environment, this can be exactly computed via the Kalman recursions.[13] In a nonlinear setting, however, computing $p(s_t|\mathcal{Z}^i_t)$ is not feasible as the density cannot be obtained analytically.

In light of this, forecasters must approximate the nonlinear state. I assume that this is done by selecting from a set of costly approximation functions ($A \in \mathcal{A}$). Forecasters first select an approximation function so as to obtain an estimate of the posterior density of the underlying state. Forecasters then report their predictions (i.e. first moment of this approximated density). Hence, the forecaster's loss function can be defined as

$$\mathcal{L} = \min_{A\in\mathcal{A}} \left\{ (\mathbf{z^i_{t+h}} - \widehat{\mathbf{z}}^{\mathbf{i,A}}_{\mathbf{t+h|t}})^\top (\mathbf{z^i_{t+h}} - \widehat{\mathbf{z}}^{\mathbf{i,A}}_{\mathbf{t+h|t}}) + c^i_A \right\} \tag{9}$$

where the first term is the mean square error arising from individual i's forecast which makes use of approximation function A, and the second term denotes the cost associated with adopting approximation function A.[14] In order to obtain heterogeneity in adoption of approximation functions, I assume that these costs are drawn randomly $c^i_A \sim U(0, \bar{c}_A)$.[15]

4.2 Approximate Predictions

After applying either of the approximation methods discussed, forecasters generate a prediction of the state and an update according to the new information received. Since agents are formulating a forecast subject to an approximation of the state, I call these approximate predictions. An approximate prediction is defined as

$$\widehat{\overline{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t}} = \int \overline{\mathbf{s}}_{\mathbf{t}}\ \widehat{p}(\overline{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}^i_t)\ d\overline{\mathbf{s}}_{\mathbf{t}} \tag{10}$$

In essence, the forecaster predicts the current state according to the approximated density $\widehat{p}(\overline{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}^i_t)$. In a linear Gaussian setting, the density is obtained exactly so that $\widehat{p}(\overline{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}^i_t) = p(\overline{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}^i_t)$ and errors are orthogonal.

Moreover, from definition (10), one can express the approximate prediction as a deviation from the optimal minimum mean square error forecast. More specifically, the approximate prediction of a model with a

[13] $p(s_t|\mathcal{Z}^i_t) \sim \mathcal{N}(s^i_{t|t}, \Psi^i_{t|t})$, where $s^i_{t|t}$ is the expected value of the posterior density and $\Psi^i_{t|t}$ is the variance.

[14] Forecasters have knowledge of the mean square error associated with each A.

[15] One could alternatively assume heterogeneous signal precision. Models featuring heterogeneous signal-to-noise ratios have been propoosed in the literature, particularly to explain forecast disagreement.

nonlinear state evolution and a linear Gaussian measurement can be expressed as

$$\widehat{\overline{\mathbf{s}}}^{i}_{t|t} = \underbrace{\mathbb{E}(\overline{\mathbf{s}}_t|\mathcal{Z}^i_t)}_{\text{Optimal}} + \underbrace{\int \overline{\mathbf{s}}_t[\widehat{\mathbf{p}}(\overline{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t}) - \mathbf{p}(\overline{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t})]d\overline{\mathbf{s}}_\mathbf{t}}_{\text{Approximation error}} \tag{11}$$

To understand how approximation errors might generate over- and underreactions to different macroeconomic variables, note that there is no restriction imposed on the sign of these mistakes. Furthermore, forecast error predictability is entirely explained by predictability of approximation errors. Whereas existing theories of expectation formation restrict the deviation from the optimal forecast to be either positive (overreactions) or negative (underreactions), the direction of the approximation error is unrestricted.

The following definition explicitly states how the sign of the errors-on-revisions coefficient is obtained:

Definition 4.1. From the nonlinear model subject to approximation errors, the covariance of errors and revision can be signed as follows:

$$\begin{aligned}\beta_1 \propto \mathbb{C}\Bigg(\overline{\mathbf{s}}_\mathbf{t}, \int \overline{\mathbf{s}}_\mathbf{t}[\widehat{\mathbf{p}}(\overline{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t}) - \widehat{\mathbf{p}}(\overline{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t-1})]\mathbf{d}\overline{\mathbf{s}}_\mathbf{t}\Bigg)\\ - \mathbb{C}\Bigg(\int \overline{\mathbf{s}}_\mathbf{t}\widehat{\mathbf{p}}(\overline{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t})d\overline{\mathbf{s}}_\mathbf{t}, \int \overline{\mathbf{s}}_\mathbf{t}[\widehat{\mathbf{p}}(\overline{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t}) - \widehat{\mathbf{p}}(\overline{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t-1})]\mathbf{d}\overline{\mathbf{s}}_\mathbf{t}\Bigg)\end{aligned}$$

Note that when there are no approximation errors, error orthogonality holds and $\beta_1 = 0$. In the case of non-zero approximation errors, however, the first term is the source of observed underreaction while the second term governs the extent of overreaction. It worth noting that the quality of an approximate prediction does not effect β_1 in a trivial manner. Instead, the way in which forecasters update their approximate predictions informs whether β_1 is positive or negative.

When forecast revisions are more closely related to the underlying state, then underreactions arise as the first term dominates the second. If instead, forecast revisions covary more with the current prediction than the underlying state, then overreactions result. In essence, when the approximate revision incorporates more noise than is optimally called for, then forecasters will appear to overreact.

A Simple Example

Suppose that the state is described as follows

$$s_t = \rho s_{t-1} + \sigma_t w_t, \qquad w_t \sim \mathcal{N}(0, \sigma_w^2)$$

$$\sigma_t = \chi \sigma_L + (1-\chi)\sigma_H$$

where $\chi = 1$ with probability q and $\chi = 0$ with probability $1-q$. Forecasters receive private signals $y_t^i = s_t + v_t^i$ each period, where $v_t^i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_v^2)$. Forecasters do not know the value of s_t or q, and can only observe the sequence of private signals $\{y_k^i\}_{k=0}^t$.[16]

To see how overreactions arise, suppose that forecasters assess the volatility of the state to be $\sigma_t = \frac{1}{2}(\sigma_L + \sigma_H)$, and forecast according to the Kalman filter subject to this naive approximation of the underlying state volatility. When q is closer to one, the signal-to-noise ratio is low and forecasters believe the state to be more variable than it truly is. As a result, forecasters place a larger weight on new information. This undue importance to news will generate overreactions. According to this simple example, the magnitude of the overreaction depends importantly on the signal-to-noise ratio. Noisier environments deliver more negative β_1 coefficients.

On the other hand, an underreaction could arise under slightly different dynamics. One can verify this by simply allowing q to be closer to zero so that the signal-to-noise ratio rises. Now, when forecasters generate approximations, they believe the state to be less variable than it truly is thereby placing less weight on news. This mutes the effects of signal noise and creates more inertia in expectation formation than is optimal. The result is, $\beta_1 > 0$.

Though this is another very simple case, it illustrates how a different approximation strategy can deliver underreactions. In particular, this updating procedure delivers smoother revisions. Moreover, the extent of observed underreaction is increasing in the variance of the state. If forecasters could simply observe q, however, then there would be no need to approximate the variance of the state and error orthogonality would hold. Of course, a naive approximation will deliver a bias in predictions. However, when approximation functions come at a cost (be it cognitive, timing, or otherwise), forecasters may find it optimal to make use of such approximations. Figure 2 below plots simulated errors against simulated revisions for each of these

[16] The assumption that forecasters are unable to observe q is for simplicity. Alternatively, one could suppose that the probability is stochastic ($q + \epsilon$) and that forecasters cannot observe the innovation ϵ.

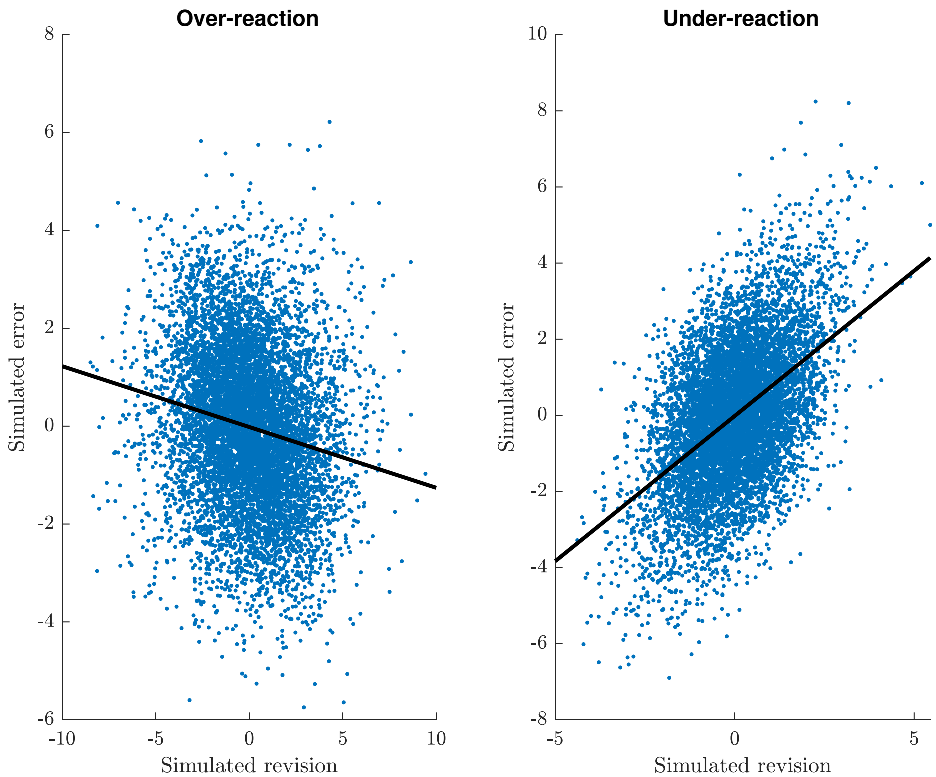


Figure 2: Simulated Examples

Note: The figures display simulated scatters plots from the examples described in the text. The panel titled “Over-reaction" assumes that $q = 0.80$ while the panel titled “Under-reaction" supposes that $q = 0.20$. The simulation is done by specifying one single forecaster for 10,000 periods, and fixing the approximation $\widehat{\sigma}_t = \frac{1}{2}(\sigma_L + \sigma_H)$

examples. The left panel shows that overreactions prevail when the revision is noisier whereas underreactions are observed when revisions are smoother.

4.3 Aggregate Error Predictability

Similar to the approximate prediction defined above, the consensus forecast arising from approximate predictions is defined as follows

Definition 4.2.

$$\alpha_1 \propto \mathbb{C}\left(\bar{\mathbf{s}}_\mathbf{t}, \int\int \bar{\mathbf{s}}_\mathbf{t}\left[\widehat{\mathbf{p}}(\bar{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t}) - \widehat{\mathbf{p}}(\bar{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t-1})\right] d\bar{\mathbf{s}}_\mathbf{t} di\right)$$
$$- \mathbb{C}\left(\int\int \bar{\mathbf{s}}_\mathbf{t}\widehat{\mathbf{p}}(\bar{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t}) d\bar{\mathbf{s}}_\mathbf{t} di, \int\int \bar{\mathbf{s}}_\mathbf{t}\left[\widehat{\mathbf{p}}(\bar{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t}) - \widehat{\mathbf{p}}(\bar{\mathbf{s}}_\mathbf{t}|\mathcal{Z}^\mathbf{i}_\mathbf{t-1})\right] d\bar{\mathbf{s}}_\mathbf{t} di\right)$$

In the previous subsection, it was determined that more volatile revisions increase the scope for overreaction. Upon aggregating (symmetrically) across several individual forecasts, the consensus revision will exhibit more persistence than the individual revisions. This motivates the next result

Proposition 2. *In the nonlinear noisy information model, $\alpha_1 \geq \beta_1$.*

Proof. See Appendix A. $\square$

The model implies that the OLS coefficient estimated from an errors on revisions regression will be weakly greater than the analogous coefficient obtained from a pooled regression of individual forecasters. This result does depend on the presence of nonlinear dynamics. In fact, this holds in the linear setting as well (see Proposition 1).[17]

4.4 Relation to Some Theories of Expectation Formation

Several other theories of expectations formation have been proposed in the literature. Here, I consider a few prominent theories and assess their ability to generate the empirical facts presented in Section 2.

Diagnostic Expectations

According to diagnostic expectations, forecasters over-weight new information according to a parameter $\theta > 0$ which governs the extent of overreactions. This parameter comes from the representativeness heuristic of Kahneman and Tversky (1974). The diagnostic nowcast is defined as $x^{i,\theta}_{t|t} = x^i_{t|t} + \theta(x^i_{t|t} - x^i_{t|t-1})$. This theory makes use of a distorted Kalman filter called the diagnostic Kalman filter. Nonetheless, as shown in Bordalo et al. (2019), diagnostic expectations are able to generate $\beta_1 < 0$ and $\alpha_1 > 0$. One shortcoming of this theory is that it is unable to explain individual underreaction ($\beta_1 > 0$). While the majority of macroeconomic time series can be characterized by overreactions at the forecaster-level, some such as unemployment, robustly exhibit underreaction at the individual level. Because $\theta > 0$, this theory cannot generate such underreaction. In other words, diagnostic expectations imply that errors and revisions always covary negatively due to the sign restriction imposed on θ.

[17] This need not be the case in other economic settings in which agents take actions, and their decisions are aggregated in a manner other than by taking a simple mean. Depending on the context, it is possible for the aggregate decision to also exhibit overreaction or excess volatility. Bordalo et al. (2019) provide a discussion of this.

Overconfidence

Models of overconfidence also distort the Kalman gain. The distortion stemming from overconfidence, however, is different. According to this theory, forecasters misperceive the precision of their own signals. Suppose that forecasters observe only one private signal, $z_t^i = s_t + v_t^i$ with $v_t^i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$. Then, forecasters perceive $\widetilde{\sigma}_v < \sigma_v$ (they believe their own private signals to be more precise than they truly are). This results in a distorted signal-to-noise ratio that forecasters use when computing their filtered estimates. Overconfident beliefs are recursive so that the distorted gain injects a bias to the update in each period. These beliefs are then projected forward only to be further distorted by the over-weighting of new information in the subsequent period. In other words, the use of $\widetilde{\kappa}$ (which is larger than the optimal gain, κ) compounds so that at an arbitrary point in time, forecasters exhibit both a non-zero ex-ante forecast error as well as a weighting error. Models of overconfidence can generate individual overreaction as well as aggregate underreaction, however, this model is similarly unable to generate individual underreaction.

Strategic Interaction

Strategic interaction models can also generate error predictability. For instance, models of strategic substitution can drive errors and revisions in opposite directions. This is because forecasters have a dual objective of not only minimizing their errors but also of distinguishing themselves from the average forecast. These models differ from the previous two in that strategic interaction models are rational. The reported forecast can also be expressed as deviation from the optimal forecast. The deviation is a function of the degree of strategic substitutability (complementarity) and the deviation of the optimal forecast from the consensus forecast. While this class of models can generate $\beta_1 > 0$ or $\beta_1 < 0$ depending on the strategic motive assumed, it is unable to jointly deliver $\beta_1 > 0$ and $\beta_1 < 0$.

Noisy Memory

More recently, a model of noisy memory was introduced in Azeredo da Silveira and Woodford (2019). In a noisy memory model, forecasters do not have access to their full history of signals due to finite memory capacity. The authors find that while rational inattention can explain individual underreactions, noisy memory may explain individual overreactions. From this model, however, it is not immediately obvious how aggregate underreactions could arise when all forecasters overreact. It also does not directly explain why the same forecaster might overreact and simultaneously underreact to different variables.[18] Lastly, the notion of forgetting past information, while compelling in general, is difficult to fathom among professional forecasters who can presumably store past forecasts and new information with relative ease.

Several other theories of expectations formation have been found to be inconsistent with the data. Models of reputational concerns imply smoothing which can only generate $\beta_1 > 0$. Moreover, asymmetric loss functions deliver counterfactually biased expectations, whereas the data show that professional forecasts are not unconditionally biased.

5 A Stylized Model of Heterogeneous Rational Approximations

To extract further insight as to how time-varying volatility can generate observed overreactions, I consider next a simple model of stochastic volatility in which forecasters select an approximation procedure. Let the underlying state be described as follows:

$$\begin{aligned} s_t &= \rho s_{t-1} + e^{h_t/2} w_t, \quad w_t \sim \mathcal{N}(0,1) \\ h_t &= \phi_0 + \phi_1 h_{t-1} + \sigma_\eta \eta_t, \quad \eta_t \sim \mathcal{N}(0,1) \end{aligned} \tag{12}$$

Furthermore, forecasters observe a contemporaneous private signal as well as a lagged public signal [19]

$$\begin{aligned} y_t^i &= s_t + v_t^i, \qquad v_t^i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma_v^2) \\ x_{t-1} &= s_{t-1} + e_{t-1}, \quad e_{t-1} \sim \mathcal{N}(0,\sigma_e^2) \end{aligned} \tag{13}$$

The specific assumptions on the volatility of s_t implies that this model is nonlinear. In addition to being unable to observe s_t, forecasters are also unable to observe h_t. As such, the innovation w_t now enters into the state multiplicatively. Moreover, let $\mathbf{z}_t^i = \begin{bmatrix} y_t^i & x_{t-1} \end{bmatrix}^\top$ denote the vector of signals observed by forecasters.

Each period consists of two stages. In the first stage, a forecaster observes $\mathbf{z_t^i}$ and selects an approximation function from a finite set $A \in \mathcal{A}$. For simplicity, define $\mathcal{A} = \{A_{KF}, A_{PF}\}$. The forecaster must pay a cost for using any A. Then in the second stage, given the approximation function and the history of signals $\mathcal{Z}_t^i$,

[18] On both of these points, developing a hybrid rational inattention-noisy memory model coud be a promising endeavor.

[19] One can alternatively envision that forecasters observe a macroeconomic variable with a transitory (e_t) and persistent (s_t) component. The persistent component is what is relevant for forecasting the target variable as it governs the long-run evolution of the random variable. Furthermore, this component is unobserved.

the forecaster reports a prediction of the public signal $\widehat{x}^i_{t|t}$.[20]

5.1 Loss Function

With this in mind, forecasters minimize the following loss function

For simplicity, I normalize $\overline{c}_{KF} = 0$. As a result, the forecaster's objective is

$$\mathcal{L} = \min\left\{MSE_{KF}, MSE_{PF} + c^i_{PF}\right\}$$

a forecaster will choose to make use of the more sophisticated PF if and only if

$$MSE_{KF} - MSE_{PF} \geq c^i_{PF} \tag{14}$$

The lefthand side of the inequality reflects the benefit to adopting the PF when formulating a prediction whereas the righthand side denotes the relative cost to adopting the PF.

5.2 Forecasting with Stochastic Volatility

This basic model belongs to the class of models described in the previous section. Importantly, this model is based on the principle that forecasters remain rational in the sense that they exhibit optimizing behavior. Clearly, $\beta_1 \neq 0$ in general. To better understand the sign of β_1, consider again the definition from Section 4.

$$\beta_1 \propto \underbrace{\mathbb{C}(s_t, \widehat{x}^i_{t|t} - \widehat{x}^i_{t|t-1})}_{\text{Underreaction}} - \underbrace{\mathbb{C}(\widehat{x}^i_{t|t}, \widehat{x}^i_{t|t} - \widehat{x}^i_{t|t-1})}_{\text{Overreaction}}$$

This is simply the definition of the covariance between forecast errors and forecast revision. The OLS coefficient is proportional to two terms, each illustrating a particular channel through which overreaction/underreaction arises. I call the first term the underreaction channel. The extent to which the forecast revision covaries with the realized state drives the scope for underreaction in the model. The second term defines the overreaction channel. The strength of overreactions ultimately comes from the extent to which it revision covary with the nowcast (above and beyond the extent to which it covaries with the state). With this in mind, it follows that β_1 depends on the underlying amount of noise in the model. Fundamentally this is a noisy information

[20]At t, forecasters make use of their full history of signals in order to formulate a state estimate, $\widehat{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}}$. The first element of the forecasted state vector is $\widehat{s}^i_{t|t}$. Based on the assumption that $x_t = s_t + e_t$, it follows that $\widehat{x}^i_{t|t} = \widehat{s}^i_{t|t}$.

Individual	Aggregate	SNR	β_1	α_1
Underreaction	Underreaction	1.427	0.12	0.46
Overreaction	Underreaction	0.433	-0.14	0.18

Table 4: Signal-to-Noise Ratio and Implied OLS Coefficients

Note: The table simulates the errors-on-revision coefficient at the forecaster-level (β_1) and the consensus-level (α_1) for two different simulated signal-to-noise ratios. I simulate a panel of 40 forecasters over 100 quarters (with an additional 100 quarter burn-in period). Each simulated forecaster generates a prediction according to the model. I then run a pooled regression for 2000 simulated panels and report the mean point estimate.

environment in which forecasters infer the state subject to private and public signals. It stands to reason that as the signal-to-noise ratio (SNR) falls, forecast revisions are driven by the noise in the system. In this case, it is *as if* forecasters report their predictions with measurement error since an upward revision in the reported forecast will mechanically result in a more negative forecast error. On the other hand, when the SNR is high, then fluctuations in the underlying state drive the forecast revisions. As a result, an upward revision delivers a more positive forecast error.

5.3 Simulation Results

Error Predictability and Signal-to-Noise Ratios

It can be shown that the model is able to generate different signs for β_1 and α_1 depending on the parameter values. The parameters defining the state and measurement are $\{\rho, \sigma_e, \sigma_v, \phi_0, \phi_1, \sigma_\eta\}$. From the intuition described above, it follows that greater signal noise variances will generate overreactions. Moreover, greater state variability reduces the scope for overreaction.

The simulation results reported in Table 4 confirm that the nonlinear model is able to jointly explain the patterns observed in the data. Consider the difference in parameter values between models of individual underreactions and models of individual overreactions. In order to flip the sign of β_1 from $\beta_1 > 0$ to $\beta_1 < 0$, I reduce the SNR from 1.43 to 0.43.[21] These simulations suggest that observed under- and overreactions at the individual level can be explained by different underlying data generating processes among otherwise rational forecasters.

I next plot the simulated β_1 distribution across high signal-to-noise ratio and low signal-to-noise ratio parameterizations. For each of 2,000 simulations, I generate a panel of 40 forecasters over 100 periods (with

[21] Although robust and reliable estimates of the SNR are missing in this literature, CG provide some estimates using cross-country data which are in line with the simulated SNR values here. In addition, these simulated SNR values are similar in magnitude to those that I estimate in Section 7.

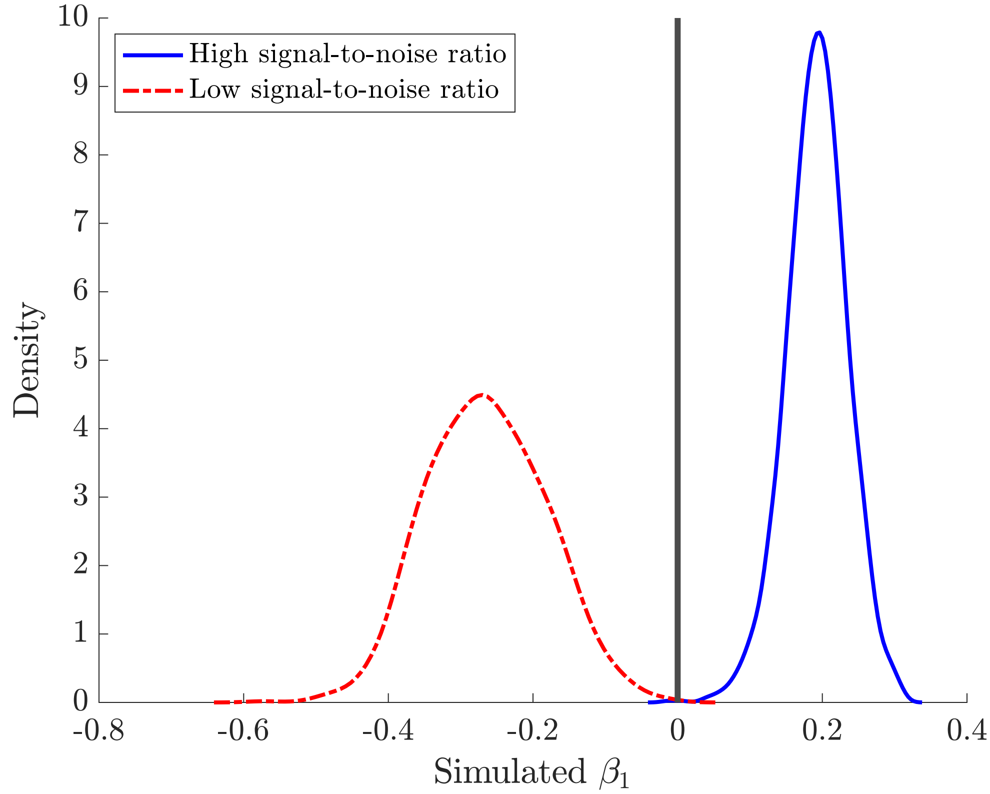


Figure 3: Covariance of Errors and Revisions Depends on Driving Process

Note: The figure plots two simulated densities of β_1 arising from a pooled individual-level errors-on-revisions regression from the model described in Section 5. The red dashed line plots the simulation in which private signal noise variance is relatively high whereas the solid blue line plots the simulation in which private signal noise variance is relatively low.

an additional 100 period burn-in). I then collect the nowcast errors and revisions for these forecasters and compute β_1. Figure 3 plots the density of β_1 across the simulations. The results confirm that over- and underreactions depend on the SNR.

6 Evidence from the Survey of Professional Forecasters

The intuition developed in Section 5 delivers model predictions that can be taken to the data. This section exploits the variation across the macroeconomic variables reported in the SPF. Each of the variables is presumed to follow a specific data generating process. As a result, $\beta_1 = \beta_1(\rho, \sigma_v, \sigma_e, \phi_0, \phi_1, \sigma_\eta)$ will, in general, vary in the cross-section of SPF variables.

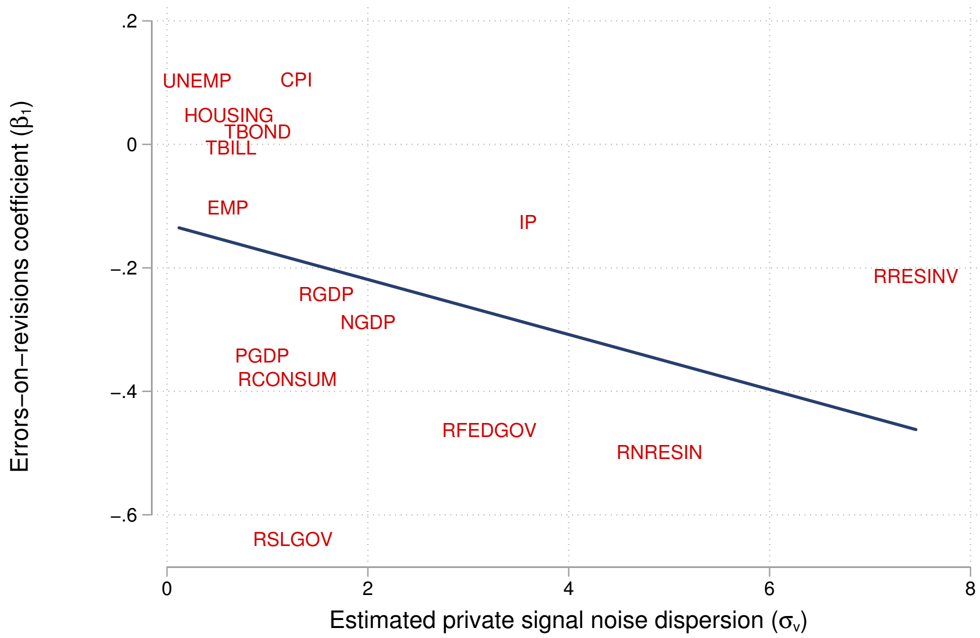


Figure 4: Error Predictability and Revision Dispersion

Note: For each SPF variable, the figure plots the estimated errors-on-revisions coefficient at the forecaster-level against estimated private noise dispersion, proxied by the interquartile range of forecast revisions. Slope of fitted line is -0.045.

Testable Prediction 1: Error Predictability and Private Noise

In this model, forecasters revise their predictions according to the realization of the lagged public signal as well as their contemporaneous private signal. The private signal noise therefore feeds into the forecast revision. From the perspective of the model, the variance of private signal noise determines the amount of dispersion in revisions across forecasters. More dispersed signal noise admits more pronounced differences in revisions.

With this insight, I collect the different β_1 coefficients across SPF variables and compute the interquartile range of revisions across forecasters for each variable. Figure 4 plots the results. As the model suggests, variables exhibiting greater dispersion in revisions tend to be those for which forecasters overreact.

Testable Implication 2: Error Predictability and Public Noise

While Figure 4 relates β_1 to private signal noise, there is also common noise present in the model. I next turn to measure the noisiness of the public signal. To reiterate, while the SPF variable of interest has sometimes been modeled as the latent state in the literature, it is best thought of as a lagged public signal. This

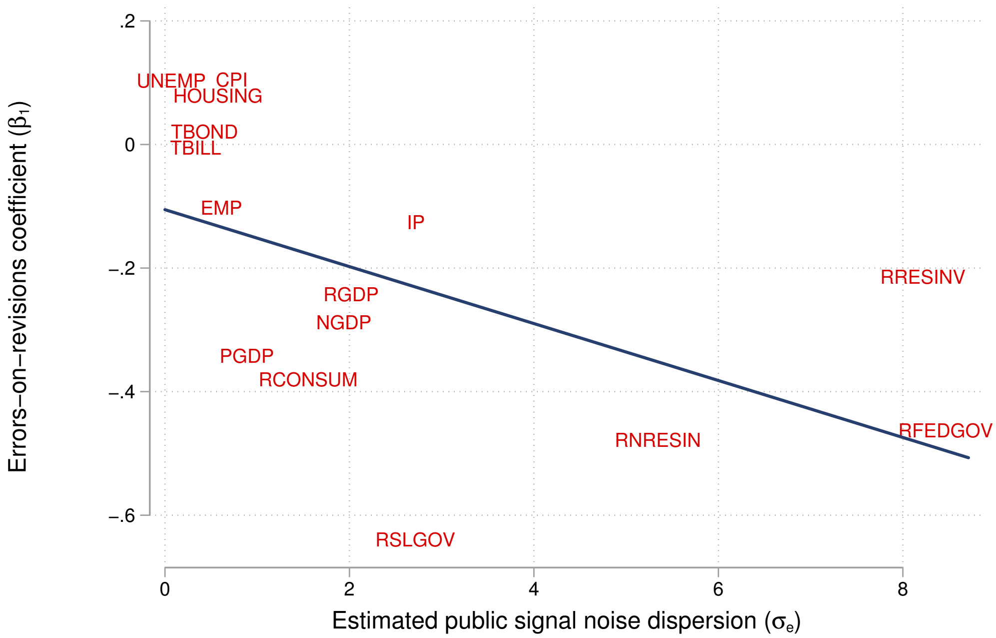


Figure 5: Error Predictability and Public Noise

Note: For each SPF variable, the figure plots the estimated errors-on-revisions coefficient at the forecaster-level against estimated public noise dispersion, proxied by the standard deviation of government revisions to real-time data. Slope of fitted line is -0.046.

is because the SPF variables are observed by all forecasters with a lag. With this in mind, the official government revisions made to these variables across different vintages can provide a measure of public signal noise. Assuming that the vintages following the initial real-time release of the variable eliminate some of the common noise, one can quantify these revisions over time. As a matter of notation, define x_t^0 as the real-time data release for a given variable, and x_t^L as the last release of the variable. Then, we can define $\text{noise}_t^{\text{public}} = x_t^0 - x_t^L$. I construct this variable from the first and last data vintage for all SPF variables in my sample, and then measure the dispersion of this public noise over time. Figure 5 relates β_1 with this measure of public signal noise. The results are consistent with the intuition of the model: variables exhibiting higher measured noise dispersion tend to deliver observed overreactions.

Testable Prediction 3: Error Predictability and Unconditional Volatility

Moreover, the model predicts that with more unconditional variability in the state, there is less scope for overreaction. To test this, I proceed to estimate ϕ_0 for each SPF variable. I then construct an estimate of

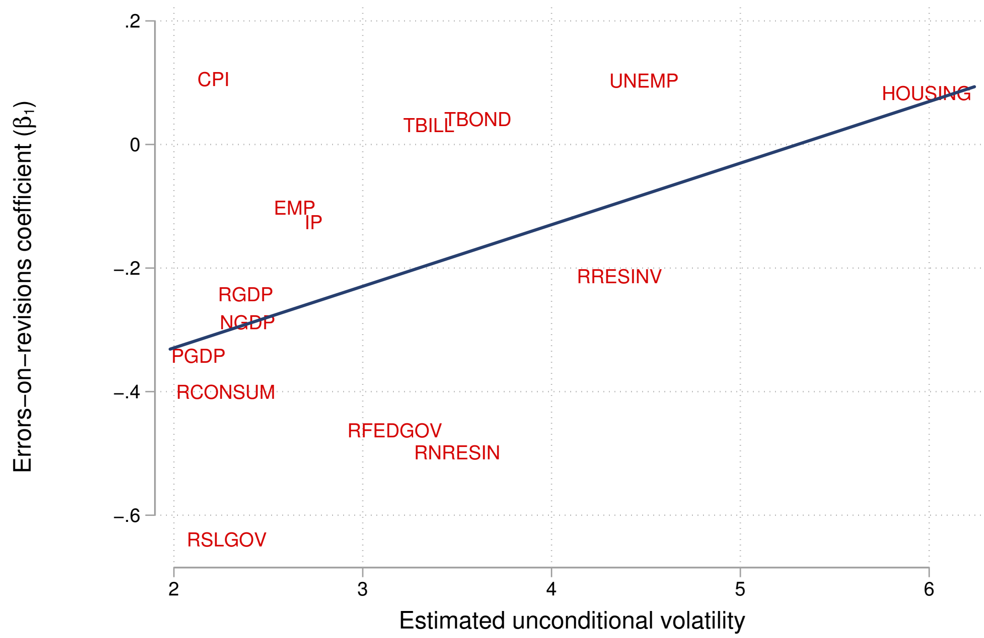


Figure 6: Error Predictability and Unconditional Volatility

Note: For each SPF variable, the figure plots the estimated errors-on-revisions coefficient at the forecaster-level against estimated unconditional volatility of the state, $\exp(\widehat{\phi}_0/2)$. Slope of fitted line is 0.100.

unconditional volatility of the state [22]

$$\text{vol}_j = \exp\left(\frac{\widehat{\phi}_0}{2}\right)$$

Figure 6 relates β_1 to vol_j. The figure supports the hypothesis that variables exhibiting more variability in the state tend to provide greater scope for underreactions. Furthermore, note that the variance of the state is increasing in ρ. Hence, the model predicts that more persistent variables will reduce the scope for overreactions. This is consistent with Bordalo et al. (2019) who verify this empirically.

Testable Prediction 4: Error Predictability and Release Frequency

As an additional way to measure signal precision, I consider the frequency with which these different variables are made available to the public. While professional forecasters report predictions in each quarter, some variables are made available at higher frequencies. Specifically the SPF conducts its survey at roughly the middle of each quarter. However, some of the SPF variables are released at a monthly frequency. For instance employment statistics are released on the Friday of each month. The survey asks forecasters to provide

[22] I estimate the parameters of the stochastic volatiliity model, $\{\phi_0, \phi_1, \sigma_\eta\}$ using MCMC techniques (see Kastner and Fruhwirth-Schattner (2014)).

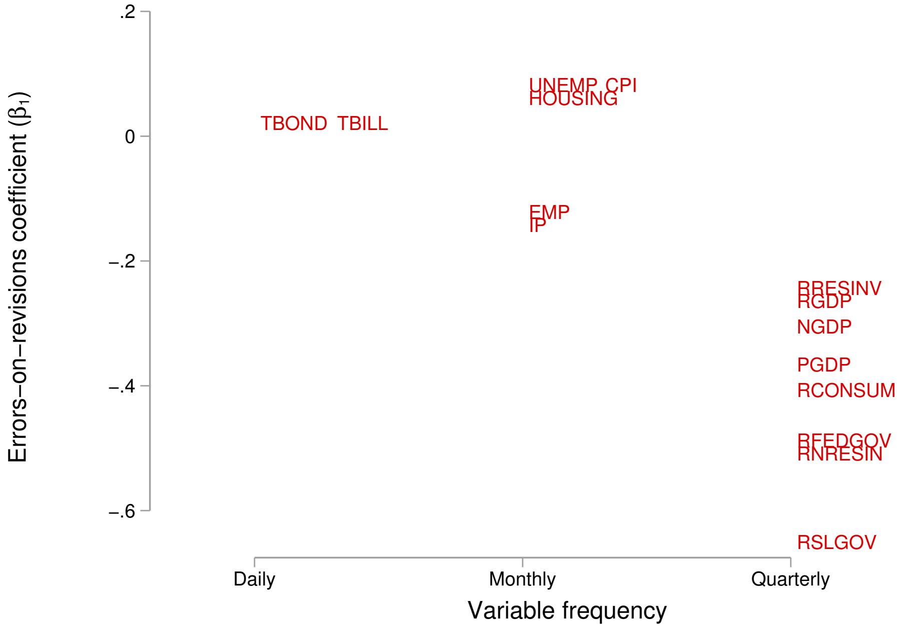


Figure 7: Error Predictability and Frequency

Note: For each SPF variable, the figure plots the estimated errors-on-revisions coefficient at the forecaster-level against variable's release frequency $\{\text{Daily}, \text{Monthly}, \text{Quarterly}\}$.

a quarterly average of these series. Furthermore, the financial time series are available at a daily frequency. As a result, forecasters have more information pertaining to some x_t at the time that they report their predictions than others. This reduces the effective noise in the lagged public signal.[23] Hence, variables available at higher frequencies should raise the scope for underreaction. Note that this does not preclude overreactive behavior in financial markets as has been readily documented. Here, I simply argue that quarterly (average of daily observations) predictions of a financial variable are better informed by the presence of daily observations through the middle of the quarter when the reported forecast is requested. On the other hand, the latest information that forecasters have for quarterly variables, such as GDP, is the previous quarter's release and an advance estimate. Since there is additional information available for some variables and not others, and the existence of this additional information depends on the variable frequency, then it follows that there is more scope for underreaction among variables that are available at higher frequencies.

[23] Alternatively, one could suppose that forecasters receive an additional informative public signal for monthly/daily SPF variables.

Jointly Testing for Overreaction and Underreaction Channels

As an additional check, I formally test for these channels jointly. For the data to accord with this theory of nonlinear expectations, it should be the case that an interaction of the forecast revision with each of these variables either raises or reduces the extent to which β_1 is negative in the pooled specification (column 1 of Table 1). To complete this exercise, I incorporate two new regressors (and all possible interactions), each capturing a source of either noise or state volatility. As a measure of noise, I select the release frequency explained above. Intuitively, SPF variables available at lower frequencies are inherently less informative and therefore raise the scope for overreaction. For my measure of fundamental volatility, I take a factor analysis approach. Since the latent state and its volatility are unobservable, it is natural to consider an index of the shared variation among all SPF variables. From this exercise, I obtain a time varying index of what I call fundamental volatility.[24]

With these two new regressors, I then modify the baseline errors-on-revisions regression (pooling across all SPF variables as in Table 1). In addition to projecting errors on revisions, I also specify the "Noise" variable (from the variable release frequency), the "Fundamental Volatility" (from the constructed index). I also include interactions of each of these with the forecast revision as well as all interactions with each other. The regression results are reported in Table 5.

The first column of the table reproduces the first column of Table 1. The second column of Table 5 reports the fully specified regression while the third column incorporates fixed effects. The relationship of interest remains the extent to which the forecast error and revision are related. Noise and fundamental volatility have no statistically detectable relation with the forecast error. In other words, these proxies cannot independently predict forecasters' mistakes. The regression also includes all possible interaction terms between the regressors. The only interaction terms that are statistically significant are those crossed with the forecast revision. Furthermore, the signs of these two interactions are consistent with the expected signs according to my theory of nonlinear expectations. In particular, noise raises the scope for overreactions as evidenced by the negative cross term between the noise proxy and the revision. On the other hand, fundamental volatility

[24] For this analysis, I drop nominal GDP since its components reside in my data set. Furthermore, I exclude CPI due to its shorter available history. For the remaining macroeconomic variables, I compute five-year rolling standard deviations and then estimate underlying principal factors. The results deliver two factors that explain roughly equal amounts of the common variance of the final vintage of SPF variables. Based on the factor loadings, I call the first factor a real residential factor, and the second a real non-residential factor (the residential factor loads highly on housing and real residential investment whereas the second factor does not). While both factors deliver the correct sign in my regression specification, I report the regression that specifies the real non-residential factor as it delivers statistically detectable results. The results are robust to 7- and 10-year window lengths, as well as specifying the real-time vintages.

	Forecast Error		
	(1)	(2)	(3)
Revision	-0.317***	-0.165**	-0.245***
	(0.050)	(0.073)	(0.056)
Revision × Noise		-0.194**	-0.115**
		(0.089)	(0.055)
Revision × Fundamental Volatility		0.106**	0.095***
		(0.041)	(0.022)
Forecaster × Variable FE	N	N	Y
Quarter FE	N	N	Y
Observations	65,070	58,740	58,740
R^2	0.06	0.07	0.16

* p<0.10, ** p<0.05, *** p<0.01

Table 5: Modified Forecast Error Predictability Panel Regressions

Note: The table reports estimated coefficients of forecast error predictability across three pooled specifications. The variable *Noise* is equal to 0 if the SPF variable is released at a non-quarterly frequency and 1 otherwise. The *Fundamental Volatility* variable is a time-varying index of fundamental volatility constructed as described in the text. Column (1) reports a simple regression of errors-on-revisions, columns (2) and (3) include the two proxies described, where column (3) includes fixed effects. In addition to the variables reported in the table, Columns (2) and (3) include the proxies individually as well as all of their interactions . Standard errors for forecaster-level regressions are as in Driscoll and Kraay (1998). Data used for estimation come from SPF (1964Q4-2018Q3).

reduces the scope for overreaction as seen by the positive coefficients in front the relevant interaction terms.

7 Parameterizing the Stylized Model

Finally, I parameterize the model in Section 5 for real GDP and unemployment. Specifically, I calibrate the public noise dispersion (σ_e), the persistence of the latent state (ρ), and the stochastic volatility parameters $(\phi_0, \phi_1, \sigma_\eta)$. I then find the values of $\overline{c}_A$ and σ_v that minimize the distance between the model-simulated and empirical estimates of $\widehat{\beta}_1$ and $\widehat{\alpha}_1$.

For real GDP and unemployment respectively, I set σ_e equal to the standard deviation of the data revisions made to each variable over the sample period. The data revision is taken to be the difference between the first and final release of the data series. For the remaining parameters, I consider the revised data rather than the real-time data. Intuitively, these series should be more highly correlated with the unobserved latent state (s_t). I then estimate an AR(1) on the revised series and set ρ equal to the estimated AR(1) coefficient. Finally, I collect the squared residuals from this autoregression and estimate $\{\phi_0, \phi_1, \sigma_\eta\}$. Table 6 reports the calibration for each variable.

Parameter	Description	Unemployment	Real GDP
ρ	State persistence	0.98	0.30
σ_e	Standard deviation of public noise	0.07	2.02
ϕ_0	Level of log variance	-0.76	0.14
ϕ_1	Persistence of log variance	0.73	0.92
σ_η	Volatility of log variance	0.69	0.39

Table 6: External Parameterization

Note: The table reports parameterization for unemployment and real GDP. Stochastic volatility parameters $\{\phi_0, \phi_1, \sigma_\eta\}$ are estimated according to the algorithm presented in Kaster and Fruhwirth-Schnatter (2014).

Finally, I parameterize σ_v and $\bar{c}_A$ by minimizing the distance between the model-implied $\{\alpha_1, \beta_1\}$ from its empirical counterpart. Computationally, this is done as follows. For each simulation, I generate a simulated state variable (s_t) according to the dynamics described in Section 5. I then simulate the lagged public signal, x_{t-1} as well as the contemporaneous private signal y_t^i. Forecasters generate a forecast in each simulated period t according to the loss function described in Section 5. From this simulated panel of forecasters, I construct the errors-on-revisions coefficients. I minimize the distance between the simulated and empirical OLS coefficients by making use of simulated annealing, a global optimization algorithm. The results are reported in Table 7.

7.1 Discussion

The results indicate that real GDP is in fact characterized by more noise than the unemployment rate. The implied signal-to-noise ratio for real GDP is about 0.58 whereas it is 0.98 for unemployment. This is consistent with the intuition of the model as well as the cross-sectional evidence in the previous section: variables

Variable	$\widehat{\sigma}_v$	$\widehat{\bar{c}_A}$	Implied SNR	Share using PF
Unemployment	0.181	0.980	0.984	0.010
Real GDP	2.092	2.831	0.583	0.685
Share that overreact to real GDP and underreact to unemployment				
Model	0.56			
Data	0.64			

Table 7: Internal Parameterization

Note: The first two columns of the top panel of the table report estimated coefficients of private signal noise dispersion and upper bound of approximation cost function. The third column reports the implied signal-to-noise ratio while the final column reports the implied share of forecasters making use of the particle filter of the naive Kalman filter. The bottom panel compares the model-implied share of over- and underreactions with the data (for real GDP and unemployment).

that exhibit higher signal-to-noise ratios are precisely those variables for which we observe underreactions.

Furthermore, the upper bound of the cost distribution is higher for real GDP than it is for unemployment. The higher upper support for real GDP is a result of the larger errors made when predicting a noisier variable. These estimates imply that nearly no forecaster opts to use the particle filter to forecast unemployment whereas almost 70% of forecasters choose the particle filter as the preferred model when predicting real GDP. This is internally consistent as real GDP is noisier than unemployment. Hence, forecasters are more likely to incur the cost to adopt a more sophisticated forecast.

Lastly, the stylized model is able to generate a surprisingly similar share of over- and underreactions. Figure 1 reports that about 64% of forecasters in my sample overreact to real GDP while simultaneously underreacting to unemployment. Based on the simple parameterization devised here, the stylized model implies that 56% of forecasters overreact to real GDP and underreact to unemployment.

7.2 Implications for Information Rigidities

What does time-varying volatility coupled with imperfect information imply about interpreting the reduced form coefficient β_1 as an information rigidity? This is a relevant question as some of the recent literature has adopted the reduced form errors-on-revisions regressions to quantiify information rigidities.[25] Based on my model, It is apparent that β_1 does not cleanly map to the Kalman gain as it does in the scalar linear context. The key intuition of Bayesian filtering, however, still holds, and the optimal weight placed on innovation errors remains a sufficient statistic for capturing the rate of learning. This weight depends on the covariances of the state estimation error and the measurement error. Quantifying the rate of learning, however, is not readily feasible from a projection of errors on revisions.

In fact, from the perspective of this model, the coefficient coming from errors on revisions regressions can reveal misleading insights on the extent of information rigidity. β_1 is positive when the SNR is high. This would typically imply larger information rigidities despite the fact that forecasters are in fact updating their beliefs more extensively. Furthermore, a negative β_1 arises when signals are less informative. For this reason, my model implies that the reduced form coefficients β_1 (and α_1) are limited in what they reveals about information frictions.

[25] See for instance Baker et al. (2020), Ryngaert (2017), Burgi (2017).

7.3 Implications for State Dependence

What does this mean for state dependence? If β_1 and α_1 were to rise in recessions, then the model would imply that the signal to noise ratio is countercyclical and information rigidities fall during economic downturns. If, on the other hand, these coefficients fall, then the signal-to-noise ratio is procyclical and information rigidities actually rise in recessions. CG document evidence indicating that α_1 falls in recessions. They interpret this as a reduction in information rigidities, however, my model would suggest that this implies a *rise* in information rigidities as the signal-to-noise ratio is falling which implies that the system experiences elevated amounts of noise (i.e. information is less precise in recessions). This is an important distinction between my model and the extant literature as it delivers an opposite answer to the question of whether individuals trust their signal more or less in recessions.

However, after performing a similar exercise to that in Table 6 by interacting a quarterly recession indicator with forecast revisions, I find no evidence that β_1 or α_1 changes with the business cycle. I also run this exercise by replacing the recession indicator with revised real GDP growth. It is possible, however, that the signal-to-noise ratio is insensitive to business cycle fluctuations because both the state and the signal experience stochastic volatility. I abstract away from volatility in signal precision, and so there is a limit to what one can glean from this model as it pertains to state dependence of information rigidities.

Nonetheless, one could distinguish between two types of uncertainty: fundamental uncertainty and information uncertainty. The first maps to time varying volatility in the state while the latter arises when signal noise experiences stochastic volatility. There is a literature that stresses the importance of uncertainty shocks. These are often modeled as fundamental uncertainty shocks, however, shocks to information precision also studied in the literature.[26] According to this model, the state dependence of β_1 and α_1 depend on the signal-to-noise ratio which in turn depends on how fundamental vs. information uncertainty evolve over the business cycle. If both rise in recessions, then is possible that the signal-to-noise ratio is acyclical thereby rendering β_1 and α_1 roughly constant over the cycle as well.

[26] For instance, Dun Jia (2016)

8 Conclusion

This paper shows that a noisy information model that incorporates time-varying volatility in the evolution of the latent state can generate error predictability at the forecaster-level. Several recent studies have explained error predictability with a model of non-rational expectations. Here, I show that this empirical fact can arise among rational forecasters attempting for predict a nonlinear state. Forecasters make unavoidable approximation errors when formulating expectations about the underlying state. These approximation errors can jointly deliver perceived under- and overreactions among forecasters. On the other hand, most behavioral theories can explain only one or the other. To this end, violations of error orthogonality are best understood as evidence against linear rational expectations.

References

[1] Andrade, Philippe and Herve Le Bihan (2013). “Inattentive Professional Forecasters" *Journal of Monetary Economics*, 60 967-982.

[2] Azeredo da Silveira, Rava and Michael Woodford (2019). “Noisy Memory and overreaction to News" *NBER Working Papers 25478*, National Bureau of Economic Research, Inc.

[3] Baker, Scott R., Tucker S. McElroy, and Xuguang S. Sheng (2020). “Expectation Formatiion Following Large Unexpected Shocks" *Review of Economics and Statistics*, 102(2) 287-303.

[4] Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2019). “Overreaction in Macroeconomic Expectations" *NBER Working Papers 24932*, National Bureau of Economic Research, Inc.

[5] Branch, William (2004). “The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations" *The Economic Journal*, 114 (July) 592-621.

[6] Brock, W.A. and Hommes C.H. (1997). “A Rational Route to Randomness" *Econometrica*, 65 1059-1160.

[7] Broer, T. and Kohlhas, A. (2018). “Forecaster (Mis-) behavior". Working Paper.

[8] Burgi, Constantin (2018). “Bias, rationality and asymmetric loss functions" *Economics Letters*, 154, 113-116

[9] Coibion, Olivier and Yuriy Gorodnichenko (2012). “What Can Survey Forecasts Tell Us About Information Rigidities?" *American Economic Review*.

[10] Coibion, Olivier and Yuriy Gorodnichenko (2015). “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts". *American Economic Review*, 109 465-490.

[11] Crisan, D. and Arnaud Doucet (2002). “A Survey of Convergence Results on Particle Filtering Methods for Practitioners" *IEEE Transactions on Signal Processing* 50 3 736-746.

[12] Crowe, Christopher (2010). “Consensus Forecasts and Inefficient Information Aggregation" *IMF Working Paper, WP/10/178*, International Monetary Fund.

[13] Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). “Investor Psychology and Security Market Under- and Overreactions." *The Journal of Finance*, 53(6), 1839-1885.

[14] Doucet, A. and Johansen A. M. (2009). “Particle Filtering and Smoothing: Fifteen Years Later" *Handbook of Nonlinear Filtering*.

[15] Dovern, Jonas amd Ulrich Fritsche and Prakash Loungani, and Natalia Tamirisa (2015). “Information Rigidities: Comparing Average and Individual Forecasts for a Large International Panel" *International Journal of Forecasting*, Elsevier 31(1), 144-154.

[16] Driscoll, John C. and Aart C. Kraay (1998). “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data" *The Review of Economics and Statistics*, 80 549-560.

[17] Dun Jia, Calvin (2016). “Disagreement vs. Uncertainty: Investment Dynamics and Business Cycles", Working Paper.

[18] Evans, G.W. and G. Ramey (1992). “Expectation Calculation and Macroeconomic Dynamics" *American Economic Review*, 82 207-224.

[19] Fuhrer, Jeff (2018). “Intrinsic Expectations Persistence: Evidence from Professional and Household Survey Expectations," Federal Reserve Bank of Boston Working Paper No. 18-9

[20] Fuster, Andreas, Benjamin Hebert, and David Laibson (2012). “Natural Expectations, Macroeconomic Dynamics, and Asset Pricing" *NBER Macroeconomics Annual* 26 Daron Acemoglu and Michael Woodford, 1-48. Chicago: University of Chicago Press.

[21] Hu, X., Thomas Schon and Lennart Ljung (2011). “A General Convergence Result for Particle Filtering" *IEEE Transactions on Signal Processing* 59 7 3424 - 3429.

[22] Julier, S. J. and Uhlman, J. K. (2004). “Unscented Filtering and Nonlinear Estimation". Proceedings of the IEEE, 92(3), 401-244.

[23] Justiniano, A. and Primiceri, G.E. “The Time-Varying Volatility of Macroeconomic Fluctuations." *American Economic Review*, 98(3), 604-641.

[24] Kahneman, Daniel, and Amos Tversky. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology 3*, no. 3 (1972): 430-454.

[25] Kim, Chang-Jin and Charles Nelson (1999). "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle" *Review of Economics and Statistics*, 81. 608-616.

[26] Kucinskas, S. and Peters, F. (2019). “Measuring Biases in Expectation Formation". Working Paper.

[27] Landier, A., Ma, Y. and Thesmar, D. (2017). “New Experimental Evidence on Expectations Formation".

[28] Lansing, K., LeRoy, S., and Ma, J. (2019). “Examining the Sources of Excess Return Predictability: Stochastic Volatility or Market Inefficiency?".

[29] Lewis, Karen(1989). “Changing Beliefs and Systematic Rational Forecast Errors with Evidence from Foreign Exchange". *The American Economic Review*, 79(4) 621-636.

[30] Mackowiak, Bartosz and Mirko Weiderholt (2009). “Optimal Sticky Prices Under Rational Inattention." *The American Economic Review*, 99 769-803.

[31] Mankiw, Gregory and Ricardo Reis (2002). “Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *The Quarterly Journal of Economics*, 117 1295-1328.

[32] McConnell Margaret, Perez-Quiros, Gabrie (2000). “Output Fluctuations in the United States: What Has Changed since the Early 1980's?" *American Economic Review*, 90(5) 1464-1476.

[33] Meinhold, Richard, J. and Nozer D. Singpurwalla (1983). “Understanding the Kalman Filter." *The American Statistician*, 37(2), 123-127.

[34] Moore, D. A. and Healy, P. J. (2008). “The Trouble with Overconfidence." *Psychological Review*, 115(2), 502.

[35] Muth, J.F.. “Rational Expectations and the Theory of Price Movements." *Econometrica*, 29(3), 315-335.

[36] Nordhaus, W. D. (1985). “Forecasting Efficiency: Concepts and Applications." Cowles Foundation Discussion papers, 774.

[37] Ottaviani, M. and Sørensen, P.N. (2006). “The Strategy of Professional Forecasting." *Journal of Financial Economics*, 81(2) 441-466.

[38] Pesaran, M.H., and Weale, M. (2006). “Survey Expectations." Working Paper.

[39] Paloviita, Maritta, and Matti Viren (2013). “Are Individual Survey Expectations Internally Consistent?" National Bank of Poland Working Paper, No. 140.

[40] Rozsypal, F. and Schlafmann, K. (2019). “Overpersistence Bias in Individual Income Expectations and its Aggergate Implications". CEPR Discussion Paper 12028.

[41] Ryngaert, Jane (2017). “What Do (and Don't) Forecasters Know About U.S. Inflation?"

[42] Sims, Christopher (2003). “The Implications of Rational Inattention." *Journal of Monetary Economics*, 50 665-690.

[43] Stock, James and Mark Watson (2007). “Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking*, 39 4-33.

[44] Stock, James and Mark Watson (2007). “Has the Business Cycle Changed and Why? (2002)" *NBER Macroeconomics Annual*, 17, 159-224.

[45] Veronesi, Pietro (1999). “Stock Market Overreaction to Bad News in Good Times: a Rational Expectations Equilibrium model" *Review of Financial Studies*, 12(5), 975-1007.

[46] Woodford, M. (2002). *Imperfect Common Knowledge and the Effects of Monetary Policy.* NBER Working Papers, Department of Economics, Columbia University.

Nonlinear Expectations: Making Sense of Professional Forecasts

APPENDIX

Julio Ortiz

Boston University

Contents

Appendix A Proofs

Proof of Proposition 1

We have the following expressions for forecast errors and revisions, respectively:

$$\mathbf{FE^i} = \mathbf{CA(I - \kappa C)(s_t - s^i_{t|t-1}) + CBw_{t+1} + Dv^i_{t+1} - CA\kappa Dv^i_t}$$

$$\mathbf{FR^i} = \mathbf{CA\kappa Dv^i_t + CA\kappa C(s_t - s^i_{t|t-1})}$$

Then,

(a) $\beta_1 \propto \text{Cov}(FE^i, FR^i) = \mathbf{CA(I - \kappa C)\Psi(CAKC)^\top - CAK(Dv^i_t v^i_t D)(CA\kappa)^\top}$

where $\mathbf{\Psi}$ denotes the state estimation error variance. This becomes

$$\begin{aligned}
\beta_1 &\propto \quad CA(I - \kappa C)\Psi(CA\kappa C)^\top - CAK(Dv^i_t v^i_t D)(CA\kappa)^\top \\
&= CA\Big\{(I - \kappa C)\Psi C - \kappa D v^i_t v^i_t D\Big\}(CA\kappa)^\top \\
\beta_1 &= 0
\end{aligned}$$

because the term in brackets is zero by the definition of the Kalman gain.

(b) Denoting $\overline{FE}$ and $\overline{FR}$ as the cross-sectional mean of the forecast error and revision, respectively, we have

$$\alpha_1 \propto \text{Cov}(\overline{\mathbf{FE}}, \overline{\mathbf{FR}}) = CA(I - \kappa C)\overline{\Psi}(CA\kappa C)^\top$$

The variance of the average revision is $\text{Var}(\overline{FR}) = CA\kappa C\Psi(CA\kappa C)^\top$ Thus, we have

$$\alpha_1 = CA(I - \kappa C)(CA\kappa C)^{-1}$$

Proof of Proposition 2

Recall that

$$\alpha_1 = [(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1})^\top(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1})]^{-1}(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1})^\top(\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})$$

and

$$\beta_1 = [(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^\top (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})]^{-1} (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^\top (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})$$

To prove the proposition, I will show that the covariance between consensus errors and revisions is weakly greater than that for pooled errors and revisions. I will then show that the variance of the consensus revision is weakly smaller than the variance of the pooled variance.

We can express the covariance between errors and revisions as

$$\begin{aligned}\mathbb{C}(\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t}, \widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) = &\int\int (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) di dt \\ &- \int\int (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t}) di dt - \int\int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) di dt\end{aligned}$$

and at the consensus level

$$\begin{aligned}\mathbb{C}(\mathbf{z}_{t+h} - \widehat{\mathbf{z}}_{t+h|t}, \widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1}) = &\int \left(\mathbf{z}_{t+h} - \int \widehat{\mathbf{z}}^i_{t+h|t} di\right)\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di\right) dt \\ &- \int \left(\mathbf{z}_{t+h} - \int \widehat{\mathbf{z}}^i_{t+h|t} di\right) dt - \int\int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1}) di dt\end{aligned}$$

We wish to show that the second equation is weakly greater than the first. One can note immediately that the second and third terms of both equations are equal (given the linearity of the expectations operator), and so they cancel out. The resulting inequality that we wish to verify is

$$\int \left(\mathbf{z}_{t+h} - \int \widehat{\mathbf{z}}^i_{t+h|t} di\right)\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di\right) dt \geq \int\int (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) di dt$$

By distributing the revision into the error on either side of the inequality, we can express each side as the sum of two terms. The first of these will drop out as we will have

$$\int \mathbf{z}_{t+h}\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di\right) dt$$

on the LHS and

$$\int\int \mathbf{z}_{t+h}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})didt$$

on the RHS. Again, due to the linearity of the expectations operator, these terms cancel out. The remaining inequality is therefore

$$\begin{aligned}
-\int\left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}]di\right)dt &\geq -\int\int \widehat{\mathbf{z}}^i_{t+h|t}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})didt \\
\int\left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}]di\right)dt &\leq \int\int \widehat{\mathbf{z}}^i_{t+h|t}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})didt \\
\int\left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)^2 dt - \int\left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)\left(\int \widehat{\mathbf{z}}^i_{t+h|t-1}di\right)dt &\leq \int\int \widehat{\mathbf{z}}^{i2}_{t+h|t}didt - \int\int \widehat{\mathbf{z}}^i_{t+h|t}\widehat{\mathbf{z}}^i_{t+h|t-1}didt \\
\int\int \widehat{\mathbf{z}}^i_{t+h|t}\widehat{\mathbf{z}}^i_{t+h|t-1}didt - \int\left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)\left(\int \widehat{\mathbf{z}}^i_{t+h|t-1}di\right)dt &\leq \int\int \widehat{\mathbf{z}}^{i2}_{t+h|t}didt - \int\left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)^2 dt \\
\int\left[\int \widehat{\mathbf{z}}^i_{t+h|t}\widehat{\mathbf{z}}^i_{t+h|t-1}di - \left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)\left(\int \widehat{\mathbf{z}}^i_{t+h|t-1}di\right)\right]dt &\leq \int\left[\int \widehat{\mathbf{z}}^{i2}_{t+h|t}di - \left(\int \widehat{\mathbf{z}}^i_{t+h|t}di\right)^2\right]dt
\end{aligned}$$

which is true since the terms in hard brackets on the RHS is the cross-sectional variance of the forecast whereas the term in hard brackets on the LHS is a cross-sectional covariance. Hence, the covariance of the consensus errors with consensus revisions is weakly greater than the covariance of individual-level pooled errors and revisions.

Finally, I show that the variance of the consensus revision is weakly smaller than the variance of the pooled revision. This is simpler to verify. Note that

$$\mathbb{V}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) = \int\int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^2 didt - \left(\int\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}]didt\right)^2$$

and

$$\mathbb{V}(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1}) = \int\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}]di\right)^2 dt - \left(\int\left[\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}]di\right]dt\right)^2$$

Once again, the second term in each of the above revision variance equations will cancel out. The resulting condition that we wish to verify is

$$\int \left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di \right)^2 dt \leq \int \int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^2 di dt$$

which holds by Jensen's inequality.

Appendix B Empirics

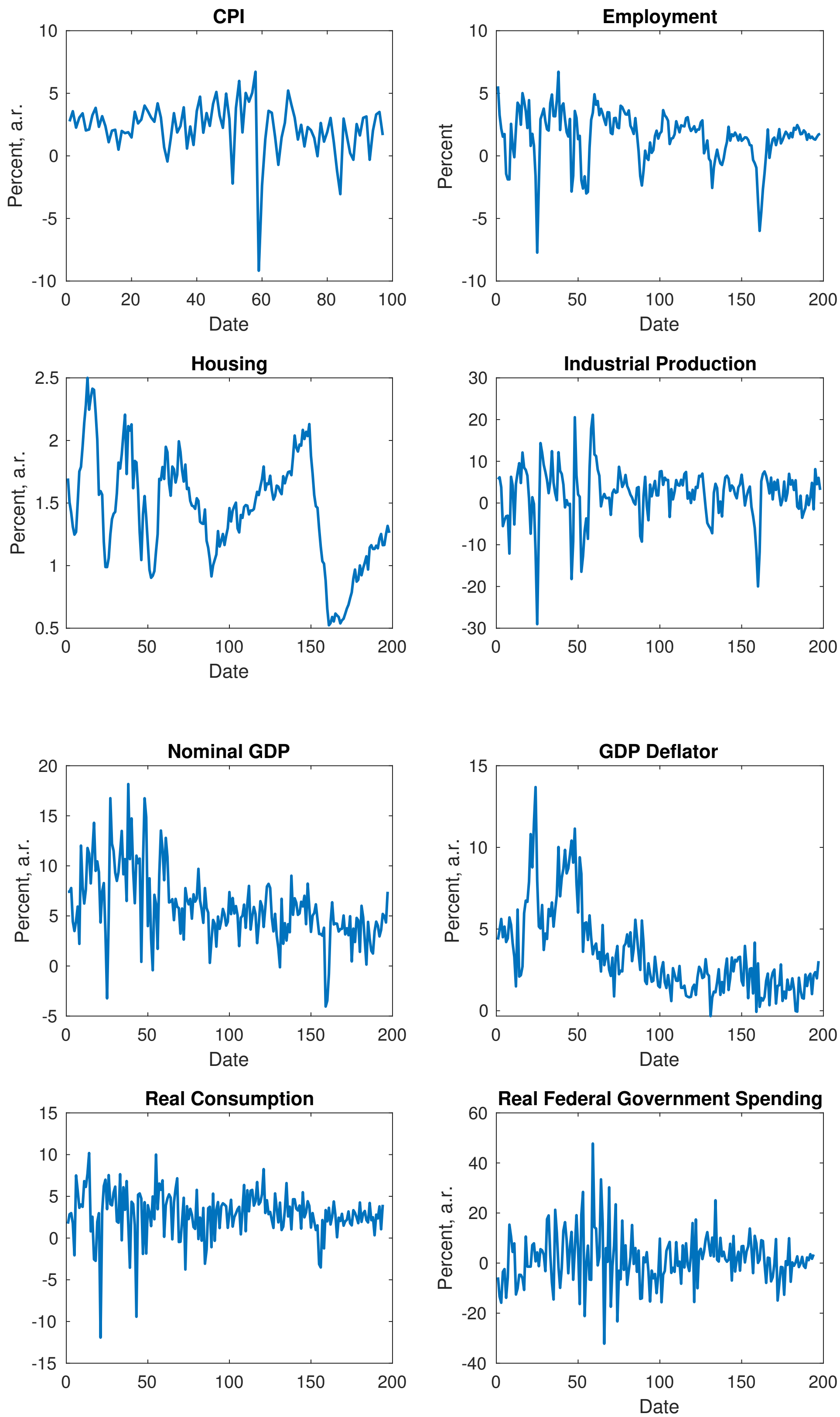
B.1 SPF: Variable Descriptions

While the paper focuses on inflation forecasts based on the GDP deflator, in this subsection I report additional results that make use of several other variables. Before presenting these results, I provide the variable descriptions below:

- NGDP–Quarterly nominal GDP growth forecast (seasonally adjusted, annual rate). Prior to 1992, these are forecasts for nominal GNP.
- RGDP–Quarterly real GDP growth forecast (seasonally adjusted, annual rate).
- PGDP–Quarterly GDP price index growth forecast (seasonally adjusted, annual rate). From 1992 - 1995, GDP implicit deflator is used, and prior to 1992, GNP implicit deflator.
- UNEMP–Forecasts for the quarterly average unemployment rate (seasonally adjusted, average of underlying monthly levels).
- EMP–Quarterly average growth of nonfarm payroll employment (seasonally adjusted, average of underlying monthly levels).
- RNRESIN–Quarterly growth forecast of real nonresidential fixed investment. Also known as business fixed investment (seasonally adjusted, annual rate).
- RRESINV–Quarterly growth forecast of real residential fixed investment (seasonally adjusted, annual rate).
- TBILL–Quarterly forecast of average three-month Treasury bill rate (percentage points, average of underlying daily levels).
- HOUSING– Quarterly growth forecast of average housing starts (seasonally adjusted, annual rate, average of underlying monthly levels).
- CPI–Quarterly forecasts of the headline CPI inflation rate (percentage points, seasonally adjusted, annual rate). Quarterly forecasts are annualized q/q percent changes of quarterly average price index level (average of underlying monthly levels).

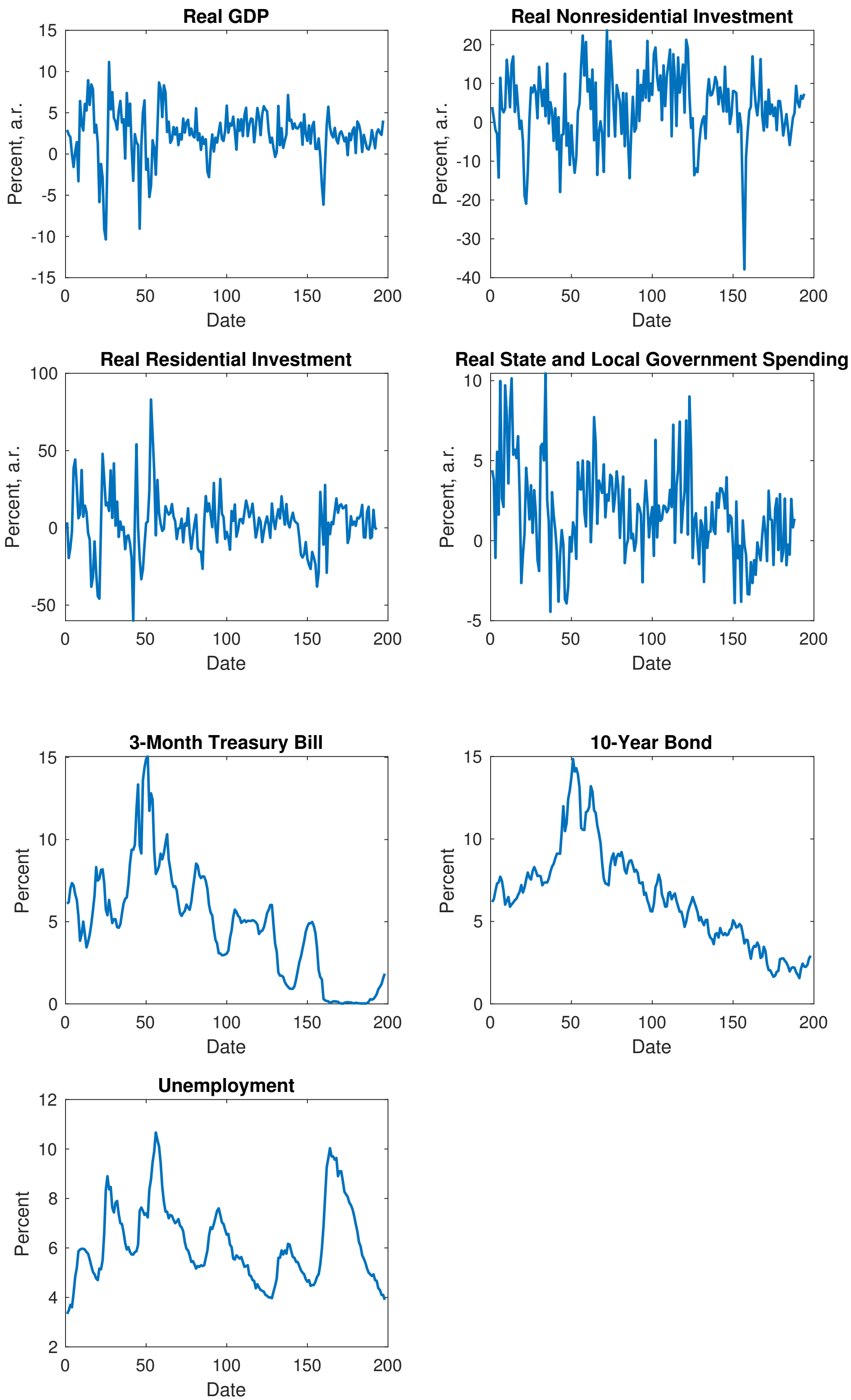
- RCONSUM – Quarterly growth forecast of real personal consumption expenditures (seasonally adjusted, annual rate).
- RFEDGOV –Quarterly growth forecast of real federal government consumption and gross investment (seasonally adjusted, annual rate).
- INDPROD – Quarterly forecasters of level of the index of industrial production, seasonally adjusted (quarterly forecasts are for quarterly average of underlying monthly levels).
- TBOND–Quarterly average 10-year Treasury bond rate (percentage points, average of the underlying daily levels). the underlying daily levels
- RSLGOV–Quarterly growth forecast of real state and local government consumption and gross investment (seasonally adjusted, annual rate).

Figure B1: Real-Time Macroeconomic Time Series



Source: Survey of Professional Forecasters

Figure B2: Real-Time Macroeconomic Time Series



Source: Survey of Professional Forecasters

B.2 Robustness: Real-Time Stochastic Volatility Regressions

Table B1: AR Specification for Volatility Regressions (Log Squared Residuals)

	AR(2)		AR(3)		AR(4)	
	ϕ_0	ϕ_1	ϕ_0	ϕ_1	ϕ_0	ϕ_1
CPI	-0.312	0.012	-0.392	0.002	-0.359	0.025
EMP	-1.129***	0.246***	-1.150***	0.236**	-1.300***	0.177*
HOUSING	-4.919***	0.149**	-4.903***	0.165**	-4.791***	0.185**
INDPROD	-1.329***	0.082	1.1389***	0.086	1.315***	0.117
NGDP	0.425**	0.116	0.377**	0.178**	0.386**	0.102
PGDP	-0.649**	0.076	-0.797***	0.096	-0.961***	0.039
RCONSUM	0.325*	0.063	0.285	-0.019	0.260	0.055
RFEDGOV	2.450***	0.217***	2.405***	0.191**	2.710***	0.109
RGDP	0.317**	0.166**	0.320**	0.153**	0.240	0.175**
RNRESIN	2.348***	0.125**	2.447***	0.106*	2.417***	0.098
RRESINV	2.846***	0.229**	2.976***	0.202**	2.694***	0.285***
RSLGOV	0.411**	0.088	0.227	0.056	0.172	0.075
TBILL	-2.129***	0.320**	-1.94***	0.381***	-1.770***	0.397***
TBOND	-2.821***	0.061	-2.760***	0.072	-2.785***	0.101
UNEMP	-3.748***	0.062	-3.972***	0.008	-3.917***	0.023

* p<0.10, ** p<0.05, *** p<0.01

Note: The table reports the estimated coefficients of an AR(1) regression on the log squared residuals of different quarterly real-time macroeconomic time series. For each variable, the log square residuals are obtained by specifying an AR(p) on the time series ($p = 2, 3, 4$). See results for $p = 1$ in main text. Newey West standard errors are specified. Data come from the Survey of Professional Forecasters (SPF).

B.3 Robustness: Simultaneous Over- and Under-Reactions

Figure 1 is computed by considering only those forecasters who contribute at least 15 forecasts to the SPF. This figure, however, is robust to the specific threshold rule used to define the sample. The following figure considers a stricter rule which requires forecasters to have contributed 30 quarters worth of forecasts. The subsequent figure eliminates this requirement entirely and considers the full sample of forecasters.

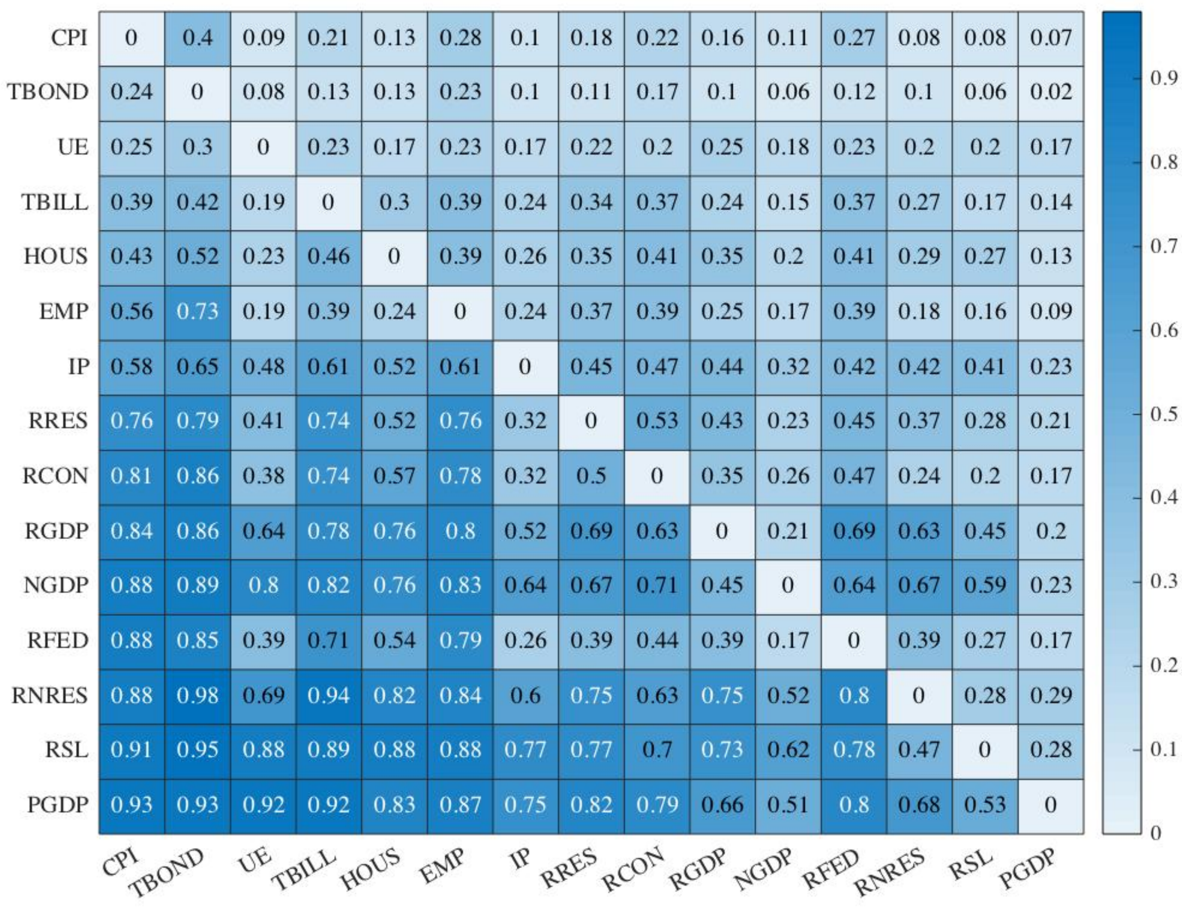


Figure B1: Frequency of Over- and Under-Reactions by the Same Forecaster

Note: This figure displays the share of forecasters who over-react to the row variable and simultaneously under-react to the column variable.

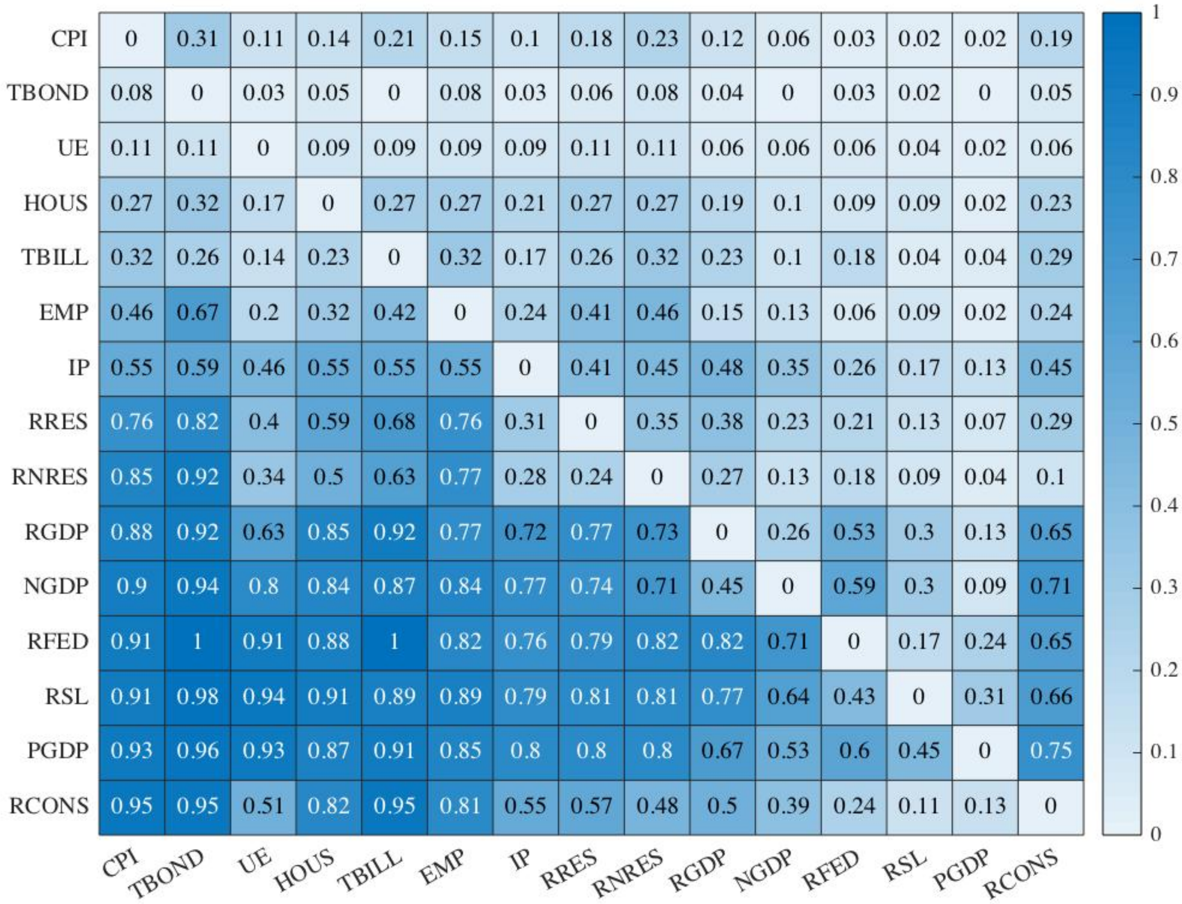


Figure B2: Frequency of Over- and Under-Reactions by the Same Forecaster

Note: This figure displays the share of forecasters who over-react to the row variable and simultaneously under-react to the column variable.

Appendix C Approximation Methods: Nonlinear Filters

C.1 Details on EKF

Consider the nonlinear model defined by (11) in the main text. The problem with invoking the standard Kalman filter is that obtaining $\mathbb{E}[F(\bar{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}_t^i)]$ is intractable. The extended Kalman filter proposes linearizing the nonlinear function $F(\cdot)$ around the optimal estimate (expanding $F(\cdot)$ about $\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}}$), to obtain

$$F(\bar{\mathbf{s}}_{\mathbf{t}}) = F(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}}) + J(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}})(\bar{\mathbf{s}}_{\mathbf{t}} - \bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}}) + h.o.t. \tag{C.1}$$

where J is the Jacobian of $F(\cdot)$ and $h.o.t.$ refers to higher order terms. The Jacobian is defined as

$$J = \begin{bmatrix} \frac{\partial F_1}{\partial \bar{\mathbf{s}}_{\mathbf{1}}} & \cdots & \frac{\partial F_1}{\partial \bar{\mathbf{s}}_{\mathbf{n}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial \bar{\mathbf{s}}_{\mathbf{1}}} & \cdots & \frac{\partial F_n}{\partial \bar{\mathbf{s}}_{\mathbf{n}}} \end{bmatrix} \tag{C.2}$$

Then,

$$F(\bar{\mathbf{s}}_{\mathbf{t}}) \approx F(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}}) + J(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}})(\mathbf{z}^{\mathbf{i}}_{\mathbf{t}} - \mathbf{z}^{\mathbf{i}}_{\mathbf{t|t-1}}) \tag{C.3}$$

therefore

$$\mathbb{E}[F(\bar{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}_t^i] \approx F(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}}) + J(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t|t}})\mathbb{E}[(\mathbf{z}^{\mathbf{i}}_{\mathbf{t}} - \mathbf{z}^{\mathbf{i}}_{\mathbf{t|t-1}}|\mathcal{Z}_t^i)] \tag{C.4}$$

Hence, the forecast is given by

$$\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t-1}} = F(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t-1|t-1}}) \tag{C.5}$$

and the update is

$$\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t}} = \widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t-1}} + \widehat{\kappa}_t(\mathbf{z}^{\mathbf{i}}_{\mathbf{t}} - \mathbf{C}\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t-1}}) \tag{C.6}$$

where $\widehat{\kappa}_t = P_{t|t-1}J^\top(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t-1}})\left[J(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t-1}})P^i_{t|t-1}J^\top(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t-1}})+\Sigma_v\right]^{-1}$ and $P^i_{t|t-1} = J(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t-1|t-1}})P_{t-1|t-1}J^\top(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t-1|t-1}})+$ Σ_w and $P^i_{t|t} = (I - \widehat{\kappa}_t J(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t|t-1}}))$

C.2 Details on UKF

The UKF makes use of the unscented transformation method which is a method for calculating moments of a random variable which undergoes a nonlinear transformation. Suppose that some random variable s (of dimension L) is propagates through a nonlinear function, $y = g(s)$. Assume that s has a mean of $\overline{s}$ and a covariance P_s. The moments of y can be computed by generating a sigma matrix χ which consists of $2L+1$ sigma vectors χ_i with associated weights W_i, according to the following:

$$\begin{aligned}
\mathcal{S}_0 &= \overline{\mathbf{s}} \\
\mathcal{S}_i &= \overline{\mathbf{s}} + \left(\sqrt{(L+\lambda)\mathbf{P_s}}\right)_i \quad i = 1, \ldots, L \\
\mathcal{S}_i &= \overline{\mathbf{s}} - \left(\sqrt{(L+\lambda)\mathbf{P_s}}\right)_{i-L} \quad i = L+1, \ldots, 2L
\end{aligned}$$

and

$$\begin{aligned}
W_0^{(m)} &= \frac{\lambda}{L+\lambda} \\
W_0^{(c)} &= \frac{\lambda}{L+\lambda} + (1-\alpha^2+\beta) \\
W_i^{(m)} &= W_i^{(c)} = \frac{1}{2(L+\lambda)} \quad i = 1, \ldots, 2L
\end{aligned}$$

where $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter.

The sigma points are propagated through the nonlinear function $\mathcal{Y}_i = g(\mathcal{S}_i)$ for all i. The mean and covariance for $\mathbf{y}_i$ is then approximated using the weights as follows

$$\begin{aligned}
\overline{\mathbf{y}}_i &\approx \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_i \\
\mathbf{P_Y} &\approx \sum_{i=0}^{2L} W_i^{(c)} \left(\mathcal{Y}_i - \overline{\mathbf{y}_i}\right)\left(\mathcal{Y}_i - \overline{\mathbf{y}_i}\right)^\top
\end{aligned}$$

The UKF makes use of this procedure by defining the augmented state vector $\begin{bmatrix} \mathbf{s_t} & \mathbf{w_t} \end{bmatrix}^\top$

C.3 Details on Particle Filtering

Rather than using generalizations to the Kalman filter, one could instead take a Monte Carlo approach. In their seminal paper, Gordon et. al. (1993) propose the bootstrap filter which is a popular variant to the particle filter. In principle, this approach makes use to mass points (particles) to approximate the underlying filtering density, $p(s_t|Z_t^i)$. This is done by defining the set of particles and associated weights: $\chi = \{s^{(n)}, \omega^{(n)}\}_{n=1}^{N}$.

Importantly, the filter still follows a general predict-update algorithm. For each particle n, the forecaster propagates the estimate through the nonlinear system

$$s_t^{i,(n)} = F\big(s_{t-1}^{i(n)}, w_t\big)$$

and then updates the weight,[1]

$$\widetilde{\omega}_t^{i,(n)} = \omega_{t-1}^{i,(n)} \cdot p\big(z_t^i | s_t^{i,(n)}\big)$$

The forecaster then normalizes the weights

$$\omega_t^{i,(n)} = \frac{\widetilde{w}_t^{i,(n)}}{\sum_{n=1}^{N} \widetilde{\omega}_t^{i,(n)}}$$

so that they sum to one. Lastly, the nowcast of the state is computed as a weighted average of the particles

$$\widehat{s}_{t|t}^i = \sum_{n=1}^{N} s_t^{i,(n)} \cdot \omega_t^{i,(n)}.$$

One common issue with sequential importance sampling is that the sample of particles tends to degenerate as few particles are given most of the weight. As a result, I make use of the common sequential importance resampling scheme in which I resample the particles, each with a probability equal to its weight.

Forecast errors and revisions are analogous to the formulation with the Kalman filter generalizations. The only difference is that the particle filtered estimates are not formulated by making use of the Kalman filtering equations. Nonetheless, these estimates approximate the optimal forecast.

[1] The precise manner in which the weights are updated depends on choices for the importance distribution.