Cautious Expectations*

Alexandre N. Kohlhas Donald Robertson

February, 2024

Abstract

This paper develops a tractable theory of cautious expectations. We impose the constraint that agents have to estimate the optimal weight on their information in an otherwise standard class of linear dynamic economies. Within this framework, we show that expectations optimally feature dampened responses to new and prior information. Our theory has several similarities to models of limited attention. However, our theory is crucially consistent with the broad-based predictability of forecast errors and biased, overreactive expectations that have otherwise called into question attention-based models. We illustrate the consequences of our framework in a standard consumption-savings problem, which shows that cautious expectations can help account for empirical evidence on the marginal propensity to consume and amplify precautionary savings.

JEL codes: C53, D83, D84, E32 *Keywords:* Expectations, information, fluctuations

1 Introduction

Uncertainty about the use and content of information plays a central role in many economic choices. Consider a household that has to decide what a raise of $x\%$ tells it about the future path of income, and hence its consumption choices. How should the household update its expectations? Building on the work of Woodford (2002) and Sims (2003), over the past two

*First draft: September 2020. Previous title: *A Rational Theory of Caution.* Kohlhas: University of Oxford (email: alexandre.kohlhas@economics.ox.ac.uk). Robertson: University of Cambridge (email: dr10011@cam.ac.uk). We are grateful to Bartosz Mackowiak, Martin Wolf, and Jieran Wu for discussing our paper at the Barcelona Summer Forum 2021, the CICM 2021 Conference, and the Nuremberg Workshop on Macroeconomic Expectations. For comments and useful suggestions, we thank Vladimir Asriyan, Pedro Bordalo, Leland Farmer, Xavier Gabaix, Zhen Huo, Per Krusell, Kristoffer Nimark, Edouard Schaal, Rosen Valchev, Mirko Wiederholt, and seminar and conference participants at the AEA Meetings 2021, 3rd Behavioral Macroeconomics Workshop, BC, CBS, ESEM 2021, ECB, Oxford University, MMF Meetings 2022, PSE, SED 2022, and IIES. This research received financial support from *Ragnar Soderberg Stiftelsen.*

decades, many advances in the theory of expectation formation have described such situations through a rational response to noisy information with known accuracy.[1] The household in our example is uncertain about what share of its raise is caused by an unobserved, persistent component, relevant to future income, and what share is related to non-persistent noise. But, because the household in this framework knows the accuracy of its information, the household crucially also knows the weight to place on current income when updating expectations.

This approach, which reduces people's uncertainty about information to one about the content of noisy signals, has proved remarkably successful. The noisy-information framework entails slow, erroneous responses of all prices and quantities, and of all agents, to all kinds of information. This contrasts with the fast, error-free reactions of the rational full-information framework and matches the real-world responses often seen in hard and survey data.[2] Yet, the success of the noisy-information framework has only been partial. Recent evidence by Bordalo *et al.* (2020) and Angeletos *et al.* (2021), amongst others, has chimed in with several earlier results to question whether rational models with noisy information can account for people's tendency to both under- and overreact to information, as well as to report biased forecasts.[3] These findings favor models which introduce further deviations to the seamless model. Despite the clear importance of this issue—additional deviations alter the outcomes and dynamics of models—no consensus has been reached about the best specification.

In this paper, we develop an estimation-based theory of expectations where people are uncertain about the accuracy of information, and hence about the *optimal weight on information*. Our framework formalizes the idea that, to form expectations, people first need to decide based on observations how much weight to accord to the various signals observed. A key difficulty faced when forming expectations lies not only in the separation of "news from noise" but also in the determination of the best use of information. The household in our example, for instance, needs to estimate the optimal weight on income when updating expectations.

Our core contribution is to propose a tractable theory of expectations that accounts for such uncertainty about the best use of information. We show that uncertainty about the optimal weight on a signal vector naturally leads to accurate-but-biased forecasts that under- and overreact to information consistent with survey data on expectations. This contrasts with other, recent models of expectation formation in which expectations either over- and underreact (e.g., Bordalo *et al.*, 2020) or are biased (e.g., Marinovic *et al.*, 2013,), but not

[1] See, for example, Sims (2003), Maćkowiak and Wiederholt (2009), Matějka (2016), Maćkowiak *et al.* (2021), and Angeletos *et al.* (2021). Earlier important contributions are Muth (1961) and Lucas (1972).

[2] See, for example, Mankiw and Reis (2002), Nimark (2008), Coibion and Gorodnichenko (2012, 2015), Angeletos and Lian (2016), Coibion *et al.* (2018), Fuhrer (2018), Kőszegi and Matějka (2020), Angeletos and Huo (2021), and Kohlhas and Walther (2021).

[3] These earlier results include, for example, Cagan (1956), Zarnowitz (1985), Cutler *et al.* (1990), De Long *et al.* (1990), Elliott *et al.* (2008), Barberis *et al.* (2018), and Bordalo *et al.* (2018).

both simultaneously.[4] We illustrate the economic consequences of our theory in a standard consumption-savings problem that lies at the cornerstone of modern macroeconomics.

To study the implications of uncertainty about the best use of information, we impose the constraint that agents have to estimate the optimal weight on observed signals in an otherwise standard class of linear-quadratic Gaussian economies. We adopt a classical view of inference so that agents estimate the optimal weight from the history of past observations alone. This avoids conflating our results with those driven by the specification of prior information. We later discuss how to incorporate prior information into our framework. Within this environment, we derive two main results.

Our first main result characterizes agents' optimal expectations. Using insights from the signal processing literature, we demonstrate that agents' expectations are characterized by a set of history-dependent *caution factors*, which scale down agents' responses to new and prior information. We describe the comparative statics of these caution factors, and discuss how they can be re-cast at the economy-wide level as a set of state-and-time dependent *attention choices*. We extend our reasoning beyond the linear-Gaussian framework: two auxiliary propositions generalize our results to prototypical examples of non-linear and non-Gaussian economies.

At the center of our first main result lies a tension between *accuracy* and *bias*. On the one hand, a decreased emphasis on any signal relative to its optimal value when agents know the accuracy of signals makes agents' expectations biased, and thus all else equal less accurate. But, on the other hand, a decreased emphasis also makes agents' expectations less volatile, all else equal improving their accuracy. We characterize this fundamental bias-variance trade-off and show that it makes a cautious response invariably optimal.

Our second main result concerns the empirical predictions of our framework. We show that our model, in which agents estimate the optimal weight on information, can match a diverse set of stylized facts on macroeconomic expectations. We proceed in two steps.

First, we document that macroeconomic expectations of output and inflation from the US Survey of Professional Forecasters (SPF) are biased but more accurate than those from popular time-series models, especially at shorter horizons. This combines the insights of Zarnowitz (1985), Elliott *et al.* (2008), and others, who show that professional forecasts are often incorrect on average, with those of Stark (2010) and Faust and Wright (2013), who show that professional forecasts often out-perform model estimates. Consistent with the predictions of our theory, we document that survey forecasts outperform model-based predictions because they are less volatile. Crucially, we show that our theory can also quantitatively match the bias-variance trade-off visible in the survey data for parameter values consistent with the SPF.

[4]See, for example, Scharfstein and Stein (1990), Bordalo *et al.* (2018), Kohlhas and Walther (2021), Angeletos *et al.* (2021), Farmer *et al.* (2022), Da Silveira *et al.* (2022), Gemmi and Valchev (2022), and Sung (2022) for both behavioral or rational models of either class.

Second, we confront our model with recent evidence on the predictability of individual forecast errors. A well-known consequence of rational (mean-squared optimal) expectations when agents know the optimal weight on information is that individual errors are unpredictable based on observed information. We show that our theory is not only consistent with previous evidence on a positive correlation between *errors* and *average revisions* (Coibion and Gorodnichenko, 2015). This evidence has lent support to noisy-information models with a known accuracy of information. But also that our model is consistent with stylized facts that are *prima facie* at odds with standard noisy-information models. In particular, we show that our model is simultaneously consistent with a negative correlation between *errors* and *individual revisions* (Bordalo *et al.*, 2020; Broer and Kohlhas, 2022). Our model of cautious responses is thus also consistent with overreactions to new information. Furthermore, our model can quantitatively account for such simultaneous under- and overeactions at the same time as matching the bias and accuracy of survey expectations.

The reason that overreactions to individual revisions arise is that agents in our framework down-weigh both new and prior information. This contrasts with standard noisy-information models, where the presence of additional noise in new information leads agents to tilt their responses away from it and towards prior information. Intuitively, when agents dampen their responses to prior information, their forecasts place only a small weight on its moderating force. Hence, when new information is positive, agents, on average, revise up their expectations by more than an agent who knows the optimal weight on information. This leads to a seeming overreaction to new information, which manifests itself in a negative correlation between errors and individual revisions. In this sense, overreactions arise in our framework because an *overreaction* to new information can be interpreted as an *underreaction* (i.e., an extra down-weighing) of prior information relative to new information.

We show that an additional testable implication of our framework is that forecasters who are more experienced should have more accurate and less biased expectations, and that forecasts accuracy should decrease by more than one-for-one in the standard deviation of shocks. Consistent with the comparative statics of our theory, we show that forecasters who are in the SPF for longer produce more accurate, less biased forecasts, and that measured increases in shock volatility are associated with a greater than one-for-one decrease in accuracy.

To explore the economic consequences of our framework, we embed our model into a multi-period consumption-savings problem with stochastic income and time-separable preferences. We choose this application because consumption and saving choices are the bedrock of modern macroeconomics. We analyze an environment in which an agent with quadratic preferences is uncertain about the best use of knowledge about her own productivity to predict future income, and show that the agent's problem maps into a special case of our framework.[5]

We document that, for standard parameters, the agent optimally chooses an upward-sloping consumption profile and muted, history-dependent responses to income shocks. The latter is caused by the agent's history-dependent dampening of expectations. The former, by contrast, is driven by the agent's caution factors depending on the number of time-series observations. All else equal, at the start of the agent's life, where she has observed few realizations of income and productivity, the agent optimally chooses to dampen income expectations more. This causes the agent to choose lower consumption initially and save more.

Lastly, we show that the comparative statics of caution factors cause a type of *precautionary savings*, which is distinct from that caused by prudence of the utility specification (Kimball, 1990). Increases in the volatility of income shocks make it harder for the agent to infer the relationship between current productivity and future income. As a result, the agent becomes more cautious and decreases income expectations more. This, in turn, leads to a pronounced fall in current consumption and a precautionary rise in savings, despite linear decision rules. Our model, calibrated to match data on US income, argues that the empirical size of this effect is equivalent to that which would have occurred with CRRA preferences with a degree of relative risk-aversion of around seven, assuming a known optimal use of information.

Finally, two wider implications of our analysis are worth noting. First, our analysis is in spirit close to those in the rational and behavioral inattention literatures (Sims, 2003; Maćkowiak *et al.*, 2021; Gabaix, 2017), in which agents observe noisy signals with known accuracy due to limited attention. The central difference is that our theory focuses on agents' responses to uncertainty about the *use* of information. The rational and behavioral inattention literatures, by contrast, center on what information agents *choose* to observe. In this sense, our theory provides a complement rather than a substitute to those based on limited attention. Each focuses on a different "stage" of the expectation formation process.

Second, the explosion of recent empirical research questioning the full-information rational expectation paradigm, with its known optimal weight on information, has forced economists to deal with the complexity of the survey evidence on expectations. This has led to several alternatives combining multiple behavioral frictions. We view one advantage of the analysis in this paper is that it provides a step towards an integrated, data-consistent model of expectations based on a minimal set of frictions.

Related Literature: In addition to the literature cited above, this paper relates to several other areas of research. We review these in order of proximity below.

One departure from full-information and rational expectations (FIRE) that has drawn

[5] In the appendix, we further show how our framework also modulates insights about monetary policy based on the Maćkowiak and Wiederholt (2009)-environment.

considerable attention has been the learning literature following Sargent (1993). Evans and Honkapohja (2012) provide an overview. In common with this literature, we share the focus on uncertainty about structural parameters of the economy, instead of uncertainty about latent factors that are the focal point in the noisy-information literature. This also connects our work to the robust control literature (e.g., Onatski and Stock, 2000; Hansen and Sargent, 2008), the work on imperfect-optimization (e.g., Ilut and Valchev, 2022; Flynn and Sastry, 2021), and that of "Brainard uncertainty" (e.g., Brainard, 1967).[6] Rather than consider whether simple learning rules converge to the rational expectation outcome, or how policy should respond to parameter uncertainty, we instead focus on how a simple-but-general type of parameter uncertainty related to the accuracy of information can help account for survey data on expectations and explore the consumption-savings implications. In this sense, our approach is complementary to that in Farmer *et al.* (2022), who show that learning about the long-run mean of an economy can also help account for salient features of survey data on professional forecasters, such as the serial correlation of forecast errors.

The predictability of individual errors, documented in survey data, can be interpreted as a rejection of the FIRE model under the joint hypothesis that agents minimize squared forecast errors. However, as Varian (1975) and Scharfstein and Stein (1990) have argued, rational agents might choose to report forecasts that differ from their conditional expectation of a variable, due to different preferences. Our results have a similar flavor. We show that agents with mean-squared error preferences *optimally choose* to make predictable forecast errors—not because of non-quadratic preferences or behavioral biases—but because of their inherent uncertainty about the optimal weight on information. This also connects our results to a recent burgeoning literature on limited memory (e.g., Afrouzi *et al.*, 2021, Da Silveira *et al.*, 2022, and Bordalo *et al.*, 2017), where agents' may choose to make predictable forecast errors due to the the mental cost of remembering past information.

Finally, the basic trade-off that controls agents' caution factors within our framework, the trade-off between the bias and the variance of expectations, is closely related to the statistical learning literature (e.g., Hastie *et al.*, 2009; Eldar, 2008) and the literature on Bayesian shrinkage estimators (e.g., Gelman *et al.*, 2013; Canova, 2011). This line of work has mainly focused on forecasters avoiding "over-fitting" models with predictive variables by adding costs (such as $L1$ and $L2$ norms)[7] or tight priors centered around zero to the inclusion

[6] Interestingly, Brainard's contribution can be seen as a specific (constrained) signal-design problem: Brainard (1967) studies how a policymaker should optimally set a policy instrument when faced with uncertainty about its effect to hit a known target for output. Within the context of our framework, Brainard's analysis hence studies the optimal choice of "a signal" given a fixed amount of uncertainty about "a response coefficient". In this sense, our paper studies the opposite problem from that in Brainard's classical analysis.

[7] These create in linear-normal models, respectively, the *Ridge Estimator* (often referred as Tikhonov regularization; Tikhonov and Arsenin, 1977; L2 norm) or the *Lasso Estimator* (Tibshirani, 1996; L1 norm).

of additional variables.[8] Yet the inclusion of such costs or priors leads to optimal forecasts that attempt to exploit the same bias-variance trade-off that we explore below. Through this lens, our contribution is to employ this trade-off *descriptively*, to argue that caution provides an optimal response to uncertainty about the best use of information. We further show that the desire to exploit this trade-off does not rest on the presence of any cost function.

2 A Baseline Framework

We begin by introducing our economic environment with uncertainty about the optimal weight on information. We then derive agents' optimal expectations in the next section.

2.1 Actions and Payoffs

The economy is comprised of a continuum of agents of measure one. Each agent chooses a forecast $f_t y_{t+k} \in \mathbb{R}$ at time t of the random variable $y_{t+k} \in \mathbb{R}$ at time $t+k$ to maximize her payoff $\mathcal{U} \in \mathbb{R}$, which depends upon the mean-squared error of her forecast,

$$\mathcal{U} = -\frac{1}{2}\mathbb{E}\left[\left(y_{t+k} - f_t y_{t+k}\right)^2\right], \quad k \geq 1. \tag{2.1}$$

The payoff function for an individual thus exhibits a standard quadratic loss in the distance between the predicted fundamental y_{t+k} and the agent's forecast of it $f_t y_{t+k}$. Nevertheless, as we will show, several of our main results extend more broadly to the class of symmetric payoff functions, in which individuals care equally about over- and under-predictions.

The fundamental y_{t+k} is assumed to have linear conditional expectation based upon the presence of n signals x_{jt},

$$\begin{aligned} y_{t+k} &= \mathbb{E}\left[y_{t+k} \mid x_t; \beta\right] + \eta_{t+k} \\ &= \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_n x_{nt} + \eta_{t+k}, \end{aligned} \tag{2.2}$$

where $x_t \equiv \begin{bmatrix} x_{1t} & x_{2t} & \ldots & x_{nt} \end{bmatrix}'$, $\beta \equiv \begin{bmatrix} \beta_1 & \beta_2 & \ldots & \beta_n \end{bmatrix}' \in \mathbb{R}^n$, and $\eta_{t+k} \sim \mathcal{N}(0, \sigma^2)$ is white noise across time conditional on past and future x_t. The linear-normal conditional expectation assumption used in (2.2) is standard in the applied literature. For example,

[8] This line of work, in turn, builds on the influential work of Stein (1956), who first showed that with $p \geq 3$ independent normally distributed variables shrinking their associated maximum likelihood estimators for the means towards some constant decreases the sum of the mean-squared errors of the individual parameter estimates. Unlike Stein's analysis, but similar to results in the statistical learning literature, we show that shrinkage towards zero (rather than some other constant) is optimal in our framework. Unlike the statistical learning literature, the optimal shrinkage that we compute does not derive from the presence of a cost function attached to the size or number of parameters. We further discuss the relationship between our results and those explored by the statistical learning literature in Section 3.

combined with Gaussianity of the signals x_{jt}, (2.1) and (2.2) capture a commonly studied class of dynamic tracking-problems (e.g., Sims, 2003; Sims, 2006; Maćkowiak and Wiederholt, 2009), in which $\left\{y_{t+k}, \{x_{jt}\}_j\right\}_t$ evolves across time and expectations are produced in accordance with the Kalman filter. In this case, one of the predictive signals in (2.2) corresponds to agents' prior expectation (e.g., $x_{1t} = f_{t-1}y_{t+k}$). We return to this application later in Section 4.

2.2 Information Structure

We assume that all agents observe the finite history of all signals, and write $v^t = \{v_s\}_{s=1}^{s=t}$ for the history of any stochastic process v_t up to date t. Agents' information set at time t is $\Omega_t = \{y^t, x_1^t, x_2^t, \ldots, x_n^t\}$. However, unlike in workhorse (noisy) rational expectations models, we do not assume agents know the mapping between the signals they observe x_t and the conditional expectation $\mathbb{E}\left[y_{t+k} \mid x_t; \beta\right]$. Instead, the β-coefficients in (2.2) must be estimated from the data. Consequently, agents are not endowed with more information than an econometrician, but must like an econometrician estimate any predictive relationship. We refer to this assumption as the *agent-econometrican assumption,* and it provides the central feature that differentiates our environment from workhorse models of expectations.

2.3 Discussion of Environment

When agents know the mapping between the signals they observe x_t and the predicted fundamental y_{t+k}, constructing optimal forecasts is simple: the conditional expectation directly provides the utility maximizing (mean-squared optimal) forecast of y_{t+k},[9]

$$f_t y_{t+k} = \mathbb{E}\left[y_{t+k} \mid x_t; \beta\right] = x_t'\beta. \tag{2.3}$$

That agents within our framework have to estimate any predictive relationship based on a finite history of signals should, nevertheless, not be controversial. It accords with ordinary experience: because of a finite number of data releases, structural breaks in the data, limited experience with the problem at hand, or limited memory, economic expectations are often based only on a finite history of observations. The agent-econometrician assumption brings forward this feature. It is further particularly appealing for two reasons.

First, as we will show, it can account for a wide range of observations about survey forecasts, using a simple mechanism. As such, it provides a parsimonious explanation for several broad aspects of survey data on expectations. Second, by exploiting ideas about biased estimation from the signal processing literature (e.g., Eldar, 2008), it arrives at predictions that

[9] Notice that we here condition the agent's expectation $\mathbb{E}\left[y_{t+k} \mid x_t; \beta\right]$ on β. This is to make clear that the agent knows the true value of $\beta \in \mathbb{R}^n$ when constructing her optimal forecast.

are independent of additional parameters to characterize individual beliefs; free parameters for which we often have little direct evidence (e.g., Carlsson and Skans, 2012; Caplin and Dean, 2015; Dewan and Neligh, 2020). In this sense, the predictions that arise from our framework are closer to being *expectation parameter free* (Sargent, 1993, 1987). We return to the advantages and disadvantages of our framework later in Section 3.

The rest of the assumptions on our environment are standard. The quadratic preferences in (2.1) preclude pre-cautionary motives in expectation formation arising from a positive third derivate of the utility function (also known as *prudence*; Kimball, 1990). They also preclude strategic complementarities in actions and expectations, such as those studied in Morris and Shin (2002) and Angeletos and Pavan (2007), among others. Both features allow us to focus on how uncertainty about the optimal weight on the signal vector x_t alone twists agents' use of information in a manner that produces cautious responses. For the same reason, we also abstract from individual-specific signals to start with. We extend our framework to account for such possibilities in Section 4. We also here note that the quadratic preferences in (2.1) are equivalent to minimizing the relative-distance entropy between the agent's predicted model and the true model in (2.2) (as in, for example, Golan *et al.*, 1997).[10]

Finally, although the linear-normal framework that we employ is particularly tractable for our purpose, none of our main results rest on either the linearity of the environment or the normality of the shocks. Appendix B.2 shows that our main results extend to both non-linear and non-Gaussian economies. We also later comment on how to extend our results to circumstances in which agents are endowed with prior information about β.

3 Optimal Expectations

We proceed to characterize agents' optimal expectations. To do so, we first define our notion of a cautious response and review an important result from the signal processing literature. We then derive several important consequences of our framework.

3.1 Definition of Caution

Let $\hat{\beta} \equiv \begin{bmatrix} \hat{\beta}_1 & \hat{\beta}_2 & \dots & \hat{\beta}_n \end{bmatrix}'$ denote an agent's estimate of β. In accordance with this estimate, the agent's expectation of y_{t+k} is $f_t y_{t+k} = x_t'\hat{\beta}$. We say that an agent exhibits *caution* towards the jth signal x_{jt} if and only if the weight placed on the jth signal in her expectation $\hat{\beta}_j$ is, on average, below the optimal-but-unknown value β_j.

[10] We thank Edouard Schaal for this comment.

Definition 1. An agent exhibits *caution* to signal j iff. the weight on the jth signal x_{jt} in her expectation $f_t y_{t+k} = x_t'\hat{\beta}$ satisfies $\mathbb{E}\left[\hat{\beta}_j \mid x^t\right] = m_j(x^t)\beta_j$ when $\beta_j \neq 0$, where $m_j \in [0,1)$.[11]

The above definition is natural. When $m_j = 0$, an agent does not alter her expectation in response to a new realization of x_{jt}. The agent shows complete caution with respect to any information contained in the signal. By contrast, if $m_j = 1$, then the agent on average changes her expectation in response to x_{jt} precisely as a fully-informed rational agent, who knows the optimal weight on the signal, would. In this sense, m_j captures how much weight the agent places on the jth signal relative to the informed choice. We therefore call m_j the *caution* dedicated to the jth signal. Finally, note that our definition of caution is closely related to common definitions of *limited attention* (e.g., Gabaix, 2017). In both cases, agents shrink their responses to new information towards zero by a factor $m_j \in [0,1)$. The key difference is that Definition 1 allows m_j to depend on the history of signal realizations x^t, and that agents are also allowed to shrink $m_j(x^t) < 1$ their responses to signals that summarize prior information.

3.2 The Benefit from Caution

An important feature of our environment is that caution is optimal. To demonstrate why caution leads to more accurate expectations, consider an agent's payoffs in (2.1). Deducting and adding $\mathbb{E}\left[y_{t+k} \mid x_t; \beta\right]$ within the quadratic term shows that[12]

$$\mathcal{U} = -\frac{1}{2}\left(\mathbb{V}\left[\eta_{t+k}\right] + \mathbb{E}_x\left[x_t' mse\left(\hat{\beta} \mid x^t\right) x_t\right]\right), \tag{3.1}$$

where

$$\begin{aligned} mse\left(\hat{\beta} \mid x^t\right) &\equiv \mathbb{E}\left[\left(\beta - \hat{\beta}\right)\left(\beta - \hat{\beta}\right)' \mid x^t\right] \\ &= \mathbb{V}\left[\hat{\beta} \mid x^t\right] + \mathbb{E}\left[\left(\beta - \mathbb{E}\left[\hat{\beta} \mid x^t\right]\right)\left(\beta - \mathbb{E}\left[\hat{\beta} \mid x^t\right]\right)' \mid x^t\right] \end{aligned} \tag{3.2}$$

denotes the *mean-squared error matrix* of the agent's estimate of β. The first term in (3.1) illustrates the welfare loss that arises in the baseline case, where the agent knows the optimal weight on the signal vector. The second term in (3.1), by contrast, shows the added welfare loss that arises from the agent's uncertainty about the true value of β.

Equation (3.2) decomposes this additional welfare loss into two further terms. One that

[11] Definition 1 conditions on x^t in its statement of the caution factor m_j ($\mathbb{E}\left[\hat{\beta}_j \mid x^t\right] = m_j(x^t)\beta_j$). Consistent with the statistics and econometrics literatures, this is because the signals x_t themselves are random variables and because we focus on the (average) weight $\hat{\beta}_j$ after a given history of signal realizations x^t. The average caution factor, across different histories of the signal vector x^t, is $\mathbb{E}_x\left[m_j(x^t)\right]$.

[12] With $\mathbb{E}_x\left[v\left(x^t\right)\right]$ we denote the expectation of the random variable $v\left(x^t\right)$ taken with respect to x^t.

reflects the covariance matrix of the parameter estimate $\mathbb{V}\left[\hat{\beta} \mid x^t\right]$, and one that reflects the outer-product of the bias vector

$$\mathbb{E}\left[\left(\beta - \mathbb{E}\left[\hat{\beta} \mid x^t\right]\right)\left(\beta - \mathbb{E}\left[\hat{\beta} \mid x^t\right]\right)' \mid x^t\right].$$

This is important. As is well-known, least-squares has the appealing property that among all possible *unbiased* estimators of β in (2.2), least-squares has the lowest variance. As a result, least-squares provides the best *unbiased* forecast of the fundamental y_{t+k} within our framework (e.g., Harvey, 1990). However, as the following example demonstrates, the qualifier is here important. By trading-off a lower variance versus a larger bias, it is possible to obtain a parameter estimate with a lower mean-squared error, and hence a more accurate forecast. Such an improved forecast, in turn, always features a cautious response.

Example 1. The Benefit from Caution: Let $n = 1$, so that an agent's expectation of y_{t+k} is based only on one signal. Suppose further that the agent considers estimates of the form $\hat{\beta}^\star = m\hat{\beta}^{ls}$, where $\hat{\beta}^{ls}$ denotes the least-squares estimator and $m \in \mathbb{R}$. That is, suppose she considers to scale her least-squares estimate, and hence also her expectation $f_t y_{t+k} = x_t\hat{\beta}$ by m. Inserting $\hat{\beta}^\star = m\hat{\beta}^{ls}$ into (3.2) and (3.1) shows, after a few simple derivations, that

$$\mathcal{U} = -\frac{1}{2}\left(\mathbb{V}\left[\eta_{t+k}\right] + \mathbb{E}_x\left[x_{1t}^2 mse\left(\hat{\beta} \mid x^t\right)\right]\right) \tag{3.3}$$

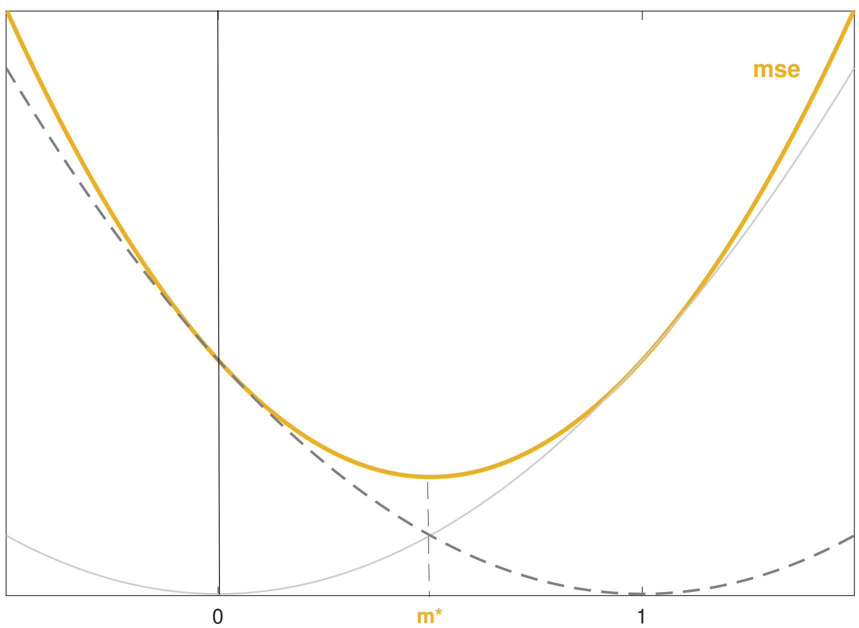
$$mse\left(\hat{\beta}^\star \mid x^t\right) = m^2\mathbb{V}\left[\hat{\beta}^{ls} \mid x^t\right] + (1-m)^2\beta^2. \tag{3.4}$$

Hence, the optimal choice of $m^\star$ for all values of x_t is always between zero and one ($m^\star \in (0,1)$; see Figure 1), and it is invariably optimal for the agent to shrink her estimate and expectation relative to least-squares. We further conclude that, as a consequence, it is optimal for the agent to exhibit caution ($\mathbb{E}\left[\hat{\beta}^\star \mid x^t\right] = m^\star\mathbb{E}\left[\hat{\beta}^{ls} \mid x^t\right] = m^\star\beta$, $m^\star \in (0,1)$).[13]

Example 1 shows that caution is optimal because it allows an agent to exploit the fundamental trade-off that exists between the variance and bias of her expectations. On the one hand, the shrunken parameter estimate associated with a cautious response has a smaller variance. This decreases the variance of the agent's forecast, all else equal improving the accuracy of her expectation. On the other hand, however, caution also leads to bias, and thus all else equal a less accurate expectation. It is this fundamental bias-variance trade-off that makes a cautious response optimal. We explore the determinants of this trade-off further in the next subsection, where we derive an agent's optimal expectation in the n-variable case and show that a simple scaling of the least-squares estimator is indeed optimal.

[13] We here exploit that the least-squares estimator is unbiased ($\mathbb{E}\left[\hat{\beta}^{ls} \mid x^t\right] = \beta$).

Figure 1: The Benefit from Caution



Note: The solid orange line in the figure depicts the mean-squared error of the parameter estimate in (3.4) for a given value of x_t. Each of the gray lines shows one of the two sub-components (either $m^2\mathbb{V}\left[\hat{\beta}^{ls} \mid x^t\right]$ or $(m-1)^2\beta^2$, respectively). The optimal value of m is denoted by $m^\star$.

3.3 Optimal Expectations

We start by characterizing a lower-bound for the variance of any estimator $\hat{\beta}$, and hence for the variance of expectations. This extends the Fréchet-Cramér-Rao (Fréchet, 1943; Cramér, 1946; Rao, 1945) lower-bound result to the general case in which estimators can be biased.

Proposition 1 (Adapted from Van Trees and Bell, 1968)**.** *Assume that the data $\{y_s, x_s\}_{s\leq t}$ follows a continuously differentiable probability density function $f(y^t, x^t; \beta)$ with respect to β, and assume that its support does not depend upon β itself. Let $\hat{\beta}$ be an estimator of β with differentiable bias function $b(\beta) \equiv \mathbb{E}\left[\hat{\beta} \mid x^t\right] - \beta$. Then, $\mathbb{V}\left[\hat{\beta} \mid x^t\right]$ is bounded below by*

$$\mathbb{V}\left[\hat{\beta} \mid x^t\right] \succeq \left[I + D(\beta)\right] J^{-1}(\beta) \left[I + D(\beta)\right]', \tag{3.5}$$

for any bias-gradient vector $D(\beta) \equiv db(\beta)/d\beta$, where

$$J(\beta) \equiv \mathbb{E}\left\{\left[\frac{d \log f(y^t, x^t; \beta)}{d\beta}\right]\left[\frac{d \log f(y^t, x^t; \beta)}{d\beta}\right]' \mid x^t\right\}$$

denotes the information matrix.[14]

To interpret Proposition 1, note that in the unbiased case, where $b(\beta) = \mathbf{0}$ and $D(\beta) = \mathbf{0}$, condition (3.5) collapses to the standard Fréchet-Cramér-Rao (FCR) bound: the minimal variance that an unbiased estimator can attain is the inverse of the information matrix. This is the variance that maximum likelihood attains, which for (2.2) is equal to least-squares. Proposition 1 states that to extend this result to a biased estimator all one needs to do is account for how the bias changes with respect to the underlying parameters. In particular, the biased FCR lower-bound depends on the *bias gradient* $D(\beta)$ and not the *bias* $b(\beta)$ itself. This makes intuitive sense since any constant bias is removable, even if it is very large, and therefore should not affect the variance of the estimator. Finally, notice that Proposition 1 extends much beyond the linear-normal case that we consider, to any continuously-differentiable density function for the data. This is important for our extensions in Appendix B.2.

A natural question that Proposition 1 raises is whether an estimator exists that attains the biased FCR lower-bound within our framework. The following proposition shows that a simple scaling of least-squares $\hat{\beta}^{ls}$ achieves this aim.[15]

Proposition 2. *Suppose $\hat{\beta} = M\hat{\beta}^{ls}$, where $M \neq I_n$ denotes an $n \times n$ matrix that can be a function of x^t. Then $\hat{\beta}$ attains the biased FCR lower-bound with bias $(M - I_n)\beta \neq 0$.*

[14] We here use the notation $A \succeq B$ to mean that $A - B$ is a positive semi-definite matrix.

[15] This proposition has appeared in several places in the signal processing literature (see e.g., Eldar, 2008).

Intuitively, Proposition 2 follows from two properties of least-squares. First, because the least-squares estimator of β attains the *unbiased* FCR lower-bound. Second, because $\mathbb{E}\left[\hat{\beta} \mid x^t\right] = M\mathbb{E}\left[\hat{\beta}^{ls} \mid x^t\right] = M\beta$, the bias function when $\hat{\beta} = M\hat{\beta}^{ls}$ is $b(\beta) = (M - I_N)\,\beta \neq 0$. A linear combination of least-squares estimators exhibits a linear bias function. Combined, these two properties ensure that $\hat{\beta} = M\hat{\beta}^{ls}$ attains the *biased* FCR lower-bound.

Equipped with Proposition 1 and 2, we are now ready to derive an agent's optimal expectation and in the process her implied caution choices. To do so, it is instructive to first re-write the agent's problem. Inserting (3.5) from Proposition 1 into the agent's payoff in (3.1) and (3.2), and using Proposition 2, shows that we can re-write her problem as

$$\max_{M \in \mathbb{R}^N \times \mathbb{R}^N} -\frac{1}{2}\left(\mathbb{V}\left[\eta_{t+k}\right] + \mathbb{E}_x x_t' \left[M J^{-1} M' + (I_n - M)\beta\beta'(I_n - M)'\right] x_t\right), \tag{3.6}$$

where the information matrix equals $J = \sigma^{-2}\sum_{\tau=1}^{t} x_\tau x_\tau'$.

The solution to this problem for all x_t provides us with Proposition 3.

Proposition 3. *Consider an agent's optimal expectation* $f_t^\star y_{t+k} = \sum_j \hat{\beta}_j^\star x_{jt} = x_t'\hat{\beta}^\star$.

(i) The optimal weight $\hat{\beta}^\star$ *on the signal vector* x_t *is*

$$\hat{\beta}^\star = M^\star \hat{\beta}^{ls}, \quad M^\star = I_n - J^{-1}\left(J^{-1} + \beta\beta'\right)^{-1} \preccurlyeq I_n, \tag{3.7}$$

where $J = \sigma^{-2}\sum_{\tau=1}^{t} x_\tau x_\tau'$ *describes the information matrix.*

(ii) The optimal weight $\hat{\beta}^\star$ *exhibits caution towards all signals,*

$$\mathbb{E}\left[\hat{\beta}_j^\star \mid x^t\right] = m^\star \beta_j, \quad m^\star\left(x^t\right) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i\beta_j\left(\frac{x_i' x_j}{\sigma^2}\right)}{1 + \sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i\beta_j\left(\frac{x_i' x_j}{\sigma^2}\right)} \in [0, 1)\,. \tag{3.8}$$

Proposition 3 characterizes an agent's optimal expectation, both in terms of the weight placed onto the signal vector $\hat{\beta}^\star$ and in terms of her implied caution choice $m^\star$. As in the simple example, the agent optimally chooses to down-weigh her responses to signal realizations, to trade-off the bias with the variance of her expectations. This is true both relative to the informed case ($m^\star < 1$), where the agent knows the optimal weight on the signal vector, as well as relative to least-squares ($M^\star \preccurlyeq I_n$). In this sense, caution provides an optimal response to the uncertainty that exists about the best use of information.

The proposition further shows that the optimal down-weighing of information depends crucially on the *information matrix* $J = \sigma^{-2}\sum_k \sum_\tau x_k x_\tau'$. This is intuitive: the more variation there is in observed signals x_t relative to noise σ^2, the easier it is for the agent to infer the optimal-but-unknown value β to attach to new information. The higher the *signal-to-noise ratio* of observations is, the less cautious the agent on average becomes. We analyze the comparative statics of agents' caution choices in detail in the next subsection, and discuss the relationship between our results and those in the statistical literature in Section 3.6.

A more subtle feature of Proposition 3 is that the agent's implied caution choice is *the same* for all signals ($m_j^\star = m^\star$ for all j). As a result, the agent in expectation shrinks her informed forecast by $m^\star$: $\mathbb{E}\left[f_t^\star y_{t+k} \mid x^t\right] = x_t'\mathbb{E}\left[\hat{\beta}^\star \mid x^t\right] = m^\star x_t'\beta$. This symmetry of caution choices arises for two reasons. First, because the agent's payoff in (2.1) depends only the properties of her forecast $f_t^\star y_{t+k} = x_t'\hat{\beta}^\star$. The expected benefit of increased caution towards one signal can therefore be mimicked by more caution towards another. Second, because the expected cost of increased caution depends only on the bias gradient $D(\beta) = db(\beta)/d\beta$ with $b(\beta) = \mathbb{E}\left[\hat{\beta} \mid x^t\right] - \beta$ and not the bias itself (see condition 3.5). Combined, these two features push the agent to optimally choose to equate the expected change in the bias gradient across different signals, and hence lead her to choose the same $m^\star$ for all x_{jt}.

This symmetry of caution choices closely resembles the symmetry of attention choices that arises in models of *rational inattention* (Sims, 2003; Maćkowiak *et al.*, 2021). A rational-inattention agent, who knows the structure of the economy, would choose to observe a univariate signal of the conditional expectation, $z_{it} = \mathbb{E}\left[y_{t+k} \mid x_t; \beta\right] + \epsilon_{it}$ with $\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Hence, as in Proposition 3, the agent would optimally choose to down-weigh all components of x_t by the same amount.[16] We further note that, as in the rational inattention literature, the optimal choice of $m^\star$ obviously depends on the structure of the variable being forecasted. If this variable is comprised of several factors $y_{t+k} = \sum_h \alpha_h y_{ht+k}$, where $\alpha_h \in \Omega_t$, with different β-coefficients on x_t, the resulting $m_j^\star$ would naturally vary across j. Maćkowiak and Wiederholt (2009) provide one example of this "different factors approach" within the rational inattention context. We consider an example within our framework in Appendix B.3.

Finally, similar to the results in the statistical learning literature (e.g., Hastie *et al.*, 2009) and Gabaix (2014), a stark feature of Proposition 3 is that the optimal response to information depends upon the true-but-unknown β. Clearly, because of the quadratic nature of (3.2), any amount of caution $m \in (2m^\star - 1, 1)$ is preferable to a full response. But, a natural question remains about how agents should best implement (3.7) and (3.8). A simple solution is to insert least-squares estimates into (3.7), and then iterate. Proposition B.1 in Appendix B.1 shows that the limit of such iterations dominates unbiased approaches. Another solution is to implement a robust (min-max) caution choice that is optimal even under the worst-case outcome for the true parameters. Proposition B.2 in Appendix B.1 shows how such robust choices can easily be made. Lastly, Appendix B.2 considers an alternative setup, where there are only two realizations of the shock η_{t+k}, and shows that in this case the optimal response

[16] For example, assume that $x_t'\beta \sim \mathcal{N}\left(0, \sigma_x^2\right)$. In this case, the rational inattention agent's expectation equals $mz_{it} = m\left(x_t'\beta\right) + m\epsilon_{it}$, where $m = \frac{\sigma_u^{-2}}{\sigma_u^{-2} + \sigma_x^{-2}}$. The agent dampens her responses to all x_{jt} equally.

is independent of the true parameters. In what follows, we will for simplicity assume that agents' expectations satisfy (3.7) and (3.8) with equality. In our quantitative application in Section 5, inserting least-squares estimates into (3.7) achieves 9/10th of the potential accuracy improvement with similar comparative statics—adjusting further for a Jensen's inequality term pushes this closer to 93 percent.[17] Yet, as we note, there are indeed several avenues for agents to implement real-time approximations of their optimal responses.[18]

We proceed with analyzing the comparative statics of an agent's caution choice, to cast more light on the forces behind agents' optimal expectations.

3.4 Comparative Statics of Caution Choices

An agent's implied caution choice $m^\star$ in (3.8) is driven by the informativeness of observations. To see this, let us first define a signal's *own* and *cross information.* We refer to the portion of the information matrix $J = \frac{1}{\sigma^2}\sum_k \sum_\tau x_k x'_\tau$ that is caused by a signal's own with-in sample variation relative to noise as its "own information". By contrast, the portion attributable to a signal's co-variation with another signal, we refer to as "cross information".

Definition 2. A signal x_{jt}'s *own information* is defined by $\sum_t \left(\frac{x_{jt}}{\sigma}\right)^2$, while the signal's *cross information* with another signal x_{ht}, $j \neq h$ is defined by $\sum_t \left(\frac{x_{ht}x_{jt}}{\sigma^2}\right)$.

Own and cross information are useful concepts because they provide a prism through which we can understand the comparative statics of agents' caution choices. Specifically, we can use these concepts to understand the effects of changes in the number of time-series observations, the number of signals observed, as well as the effects of changes to the with-in sample volatility and correlation of signals. Proposition 4 takes a first step towards this goal.

Proposition 4. *Consider an agent's implied caution choice $m^\star$.*

(i) $m^\star$ increases in signal j's own information.

Suppose further that $\beta_j\beta_h < (>)0$ for $h \neq j$. Then:

(ii) $m^\star$ decreases (increases) in the jth signal's cross information with h.

Proposition 4 formalizes the effects that own and cross information have on implied caution choices. The larger a signal's own information $\sum_t \left(\frac{x_{jt}}{\sigma}\right)^2$ is, the easier it is for an agent to

[17] The statistical learning literature often combines a similar approach with a cross-validation test, which uses out-of-sample observations to verify the choice of approximating parameter (Hastie *et al.*, 2009). One can further improve the accuracy of the approximation in our case by adjusting for a Jensen's inequality term that arises from β^2 having to be estimated with an unbiased approach in (3.7) rather than β itself.

[18] Gabaix (2014) also considers an alternative argument: Following Kahneman and Tversky (1973), Gabaix (2014) argues that the design of an optimal choice is done by "system 1" in the brain of an agent, under full information about the true parameters of the model. Actual choices are then enacted intuitively by "system 2", according to (3.7) and (3.8), not knowing the true parameters.

correctly infer the optimal-but-unknown value β_j to attach to the signal from a given history of observations. The signal-to-noise ratio of observations is, all else equal, larger. As a result, the agent optimally chooses to shrink her responses by less and becomes less cautious ($m^\star \uparrow$).

The second part of the proposition, by contrast, shows that cross information can scramble the information that agents have about a signal. For example, suppose the cross information between two signals is positive ($\sum_t \left(\frac{x_{ht}x_{jt}}{\sigma^2}\right) > 0$), but that one of the signals has a positive effect on y_{t+k} while the other has a negative effect ($\beta_j > 0$ while $\beta_h < 0$). Then, any variation in this latter signal makes it more difficult for the agent to infer the optimal weight to assign to x_{jt}. The variation in the second signal effectively scrambles the information about the first. Through this mechanism, cross information can cause caution to increase ($m^\star \downarrow$).

Proposition 5 uses the results in Proposition 4 to detail the effects of additional signals and time-series observations on an agent's caution choice.

Proposition 5. *Suppose all signals are orthogonal $\sum_t x_{ht}x_{jt} = 0$ for $h \neq j$. Then, for any x^t:*

(i) $m^\star$ increases in the number of signals n.

Suppose instead that $\beta_h^2 \sum_t x_{ht}^2/\sigma^2 + \beta_h \sum_{j\neq h} 2\beta_j \sum_t x_{ht}x_{jt} < 0$ for some signal h. Then:

(ii) $m^\star$ decreases with the observation of the hth signal.

Finally, irrespective of the sample covariance $\sum_t x_{ht}x_{jt}$ between signals:

(iii) $m^\star$ increases in the number of time-series observations t.

All else equal, the more signals n an agent observes, the more own information she has to infer the optimal-but-unknown weights β to attach to new information. This explains why the number of signals decreases an agent's caution ($m^\star \uparrow$) when there is no cross information ($\sum_t x_{ht}x_{jt} = 0$ for $h \neq j$). However, Proposition 5 also shows that such effects can be overturned if the cross information scaled by the true parameters is sufficiently negative. In this case, the observation of the hth signal decreases the overall information that exists, and increases the amount of caution ($m^\star \downarrow$). Thus, depending on the relative size of own and cross information effects, caution can either *decrease* or *increase* in the number of signals.

Unlike the observation of an additional signal, observing more time-series observations however always decreases an agent's caution choice. Although cross information can mitigate the increase, observing more data points invariably increases agents' knowledge about the optimal-but-unknown weights β to attach to x_t. This explains why $m^\star$ always increases in the number of time-series observations t. Finally, we note that if $t \to \infty$, $m^\star \to 1$ becomes optimal for any x^t. Although the speed of convergence depends crucially on the own- and cross information structure of signals, with an infinite amount of observations, the optimal expectation in (3.7) eventually converges to $\mathbb{E}\left[y_{t+k} \mid x_t; \beta\right]$ in (2.3).[19]

[19] We here also use that $\hat{\beta}^{ls}$ converges in probability to β as $t \to \infty$. We discuss the speed of convergence in

3.5 A Partial Equivalence Result

The optimal expectation in (3.7) resembles but is distinct from that which arises in noisy-information models where agents know the optimal weight on information. Proposition 6 explicitly compares an agent's expectation to those that arise from a simple model with noisy information. The proposition formalizes our earlier discussion by providing a partial equivalence result between the two approaches. Notice that the noisy signal we consider in the proposition corresponds to the optimal rational inattention signal (Section 3.3).

Proposition 6. *Consider a continuum of agents $i \in [0,1]$. Suppose agent i's forecast equals her conditional expectation $f_{it}y_{t+k} \equiv \mathbb{E}[y_{t+k} \mid z_{it}; m]$ based on the noisy signal*

$$z_{it} = x_t'\beta + \epsilon_{it}, \tag{3.9}$$

where $x_t'\beta \sim \mathcal{N}(0, \sigma_x^2)$ and $\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with $\mathbb{E}[\epsilon_{it}\epsilon_{\ell t}] = 0$ for $\ell \neq i$. We denote the weight on z_{it} in $\mathbb{E}[y_{t+k} \mid z_{it}; m]$ by $m(\sigma_x, \sigma_\epsilon) \in (0,1)$. Then, there exists a noise choice σ_ϵ such that

$$\int_0^1 \mathbb{E}[y_{t+k} \mid z_{it}; m]\, di = \mathbb{E}[f_t^\star y_{t+k} \mid x_t] \quad \forall x_t. \tag{3.10}$$

Proposition 6 highlights similarities and differences between the two frameworks. The proposition shows that the dampened expectations that arise at the aggregate level in the noisy-information framework resemble those which occur due to uncertainty about the best use of information. This is because, in both cases, agents shrink their expectations towards zero. However, equation (3.10) also hints at important differences.

The equation equates *individual* optimal expectations from (3.7) with *average* expectations from the noisy-information framework. The individual expectations from the noisy-information framework are more volatile, due to the reliance on the noisier signal z_{it}. Furthermore, the comparative statics of individual expectations are also different. An increase in the noise component σ_ϵ will, all else equal, cause a decline in the weight on new information m in the noisy-information framework. This will tilt agents' expectations away from new information z_{it} towards prior information $x_t'\beta \sim \mathcal{N}(0, \sigma_x^2)$. By contrast, an increase in σ in (2.2) will cause $m^\star$ in (3.8) to fall across all signals x_{jt}. This includes also signals that summarize prior information. Section 4 further explores this difference using survey data on expectations.

Finally, notice that Proposition 6 equates the average expectation from the noisy-information case with the *expected value* of the optimal expectation in (3.7) conditional on the realization of the signal vector x_t *period-by-period*. This is to account for the fact that an agent's caution choice in (3.8) is history-dependent (see also the discussion below).

Section 3.6.

3.6 Discussion and Extensions

In most economic circumstances the best means to use any piece of information is unknown. The analysis in this section has shown that such uncertainty leads agents to optimally dampen their responses to information, consistent with a cautious approach. Through this lens, we have shown that caution provides a purely rational response to the uncertainty that exists about the best weight on information. Six further features are worth highlighting.

1. *Parameter sparse expectations*: First, notice that the expectations that arise from our framework (Proposition 3) do not depend on the choice of any cost function on beliefs. In this sense, agents' expectations are closer to being "expectation parameter free" (Sargent, 1993); they do not depend on extraneous parameters to characterize frictions to individual beliefs. Our framework, as a result, provides a parsimonious model of expectations.

2. A Bayesian perspective: Second, the results that we have derived so far have all employed a classical view of inference, avoiding thorny questions about the shapes and origins of initial priors. However, our results can also be given a strictly Bayesian interpretation. Our main results (Proposition 3 to 5) are, for example, identical to those that would arise if we equip agents at $t = 0$ with a (conjugate) Gaussian prior over β with mean 0 and covariance matrix $\sigma^2\Lambda^{-1}$, where $\Lambda^{-1} \in \mathbb{R}^{n\times n}$. The mean posterior estimate of β would for any $t > 0$ always equal (3.7) for a particular choice of Λ^{-1}.[20] This shows that our results can also be interpreted within a Bayesian context. Indeed, we can in this case view the optimal expectation in (3.7) as the conditional expectation of y_{t+k} based on the history of observables (x^t, y^t) and a specific initial prior over β, $\beta \sim \mathcal{N}(0, \sigma^2\Lambda^{-1})$ ($f_t^\star y_{t+k} = \mathbb{E}[y_{t+k} \mid x^t, y^t; p_\beta]$, where p_β denotes the prior). Furthermore, for any initial prior, one can extend our results to the Bayesian case with the help of the *Bayesian FCR bound* (Aras *et al.*, 2019). Similar steps to those outlined above, replacing the FCR bound with its Bayesian counterpart that accounts for the information contained in the prior, will then characterize an agent's conditional expectation.

3. Law of Iterated Expectations: Third, the Bayesian perspective also hints at an important feature of agents' optimal expectations. As with the rational expectations framework with a known optimal use of information, the optimal expectations in (3.7) abide by a Law of Iterated Expectations: $f_{t-1}^\star f_t^\star y_{t+k} = f_{t-1}^\star y_{t+k}$ (Appendix B.4). This demonstrates that an agent with expectations characterized by Proposition 3 will never expect to revise her own expectations. Section 5 shows that a key corollary of this result is that linear dynamic models

[20] Equating the Bayesian estimator $\hat{\beta}^{\text{bayes}} = (X'X + \Lambda)^{-1} X'Y$, where $X \equiv \begin{bmatrix} x_1' & \dots & x_t' \end{bmatrix}'$ and $Y \equiv \begin{bmatrix} y_1 & \dots & y_t \end{bmatrix}'$, with $\hat{\beta}^\star$ shows that $\Lambda^{-1} = J^{-1}\sigma^{-2}\left[\left[I_n - JM^\star J^{-1}\right]^{-1} - I_n\right]$. Notice that Λ^{-1} is symmetric and positive semi-definite. The equivalence between the two estimators follows immediately in the special case in which $n = 1$; both estimators are in this case part of the "Tikhonov regularization class" (Appendix B.4).

with expectations characterized by Proposition 3 are no harder to solve than their FIRE counterpart, and can even be solved using the same solution methods.

4. Predictable forecast errors: The fourth feature is equally important. Unlike models of rational expectations with a known optimal use of information, the forecast errors that result from our framework are predictable based on information observed by individual agents. This is because each agent down-weighs her own information. Section 4 discusses the importance of this result for the ability of our framework to match recent evidence on agents' expectation formation. The predictability of individual errors, documented in the literature, can be interpreted as a rejection of the FIRE model under the joint hypothesis that agents minimize squared errors. However, as Varian (1975) and Scharfstein and Stein (1990) have argued, fully rational agents might choose to report forecasts that differ from their conditional expectation due to different preferences. Our results have a similar flavor: they show that agents with mean-squared error preferences may optimally choose to make predictable forecast errors— not because of non-quadratic preferences or any behavioral frictions— but merely because of their inherent uncertainty about the optimal weight on information.

5. History-dependent caution choices: Fifth, recall that an agent's caution choice $m^\star$ in (3.8) is history-dependent, and depends on *the realization of* x^t. This allows our framework to speak to the evidence on history-dependent expectations (Section 4). It also highlights a key difference between Proposition 3 and similar results in the behavioral and rational inattention literatures. Unlike the results in, for example, Sims (2003), how much an agent down-weighs information depends on the history of signal realizations through own- and cross-information terms, rather than only on the properties of the data-generating process. Furthermore, this history-dependence can serve as the basis for additional heterogeneity in expectations.

6. Finite sample economy: The sixth and final feature of our framework follows directly from the agent-econometrician assumption. Similar to the problem faced by econometricians, an agent in our model has only a finite sample (and hence limited information) to base her expectations on. This is what drives the agent to down-weigh information, to exploit the fundamental trade-off between the variance and the bias of her expectations. The consequences of the finite sample nature of economic observations is, in turn, exacerbated within our framework by the presence of (i) multiple structural breaks; (ii) serial correlation in the error term; (iii) a low number of degrees of freedom, due to a large number of signals n, some of which negatively co-vary; and (iv) limited memory on the part of agents. All of these features, decrease the effective informativeness of observations, and hence increase an agent's caution choice ($m^\star \downarrow$). Building on the work of Kozlowski *et al.* (2020) and others, Farmer *et al.* (2022) show that a modest increase in the complexity of the underlying model can drastically slow down learning about unknown parameters, such as the optimal weight on information. An increase in the complexity of the model in (2.2) would, all else equal, further amplify decreases in $m^\star$. Notwithstanding such additional limits to the informativeness of observations, in what follows, we continue to use a reductionist approach that caps agents' information at a fixed number of observations. This allows us, in an analytically tractable manner, to study the consequences of uncertainty about the optimal weight on information.

We close this section with one final observation. The fundamental bias-variance trade-off that is at the core of our analysis is closely related to the statistical learning literature (e.g., Hastie *et al.* 2009) as well as the literature on Bayesian shrinkage estimators (e.g., Gelman *et al.*, 2013; Canova, 2011). These lines of work have mainly focused on forecasters avoiding "over-fitting" models with predictive variables by adding costs to the inclusion of additional variables (such as $L1$ or $L2$ norms), or tight priors centered around zero. Yet the inclusion of such costs or priors leads to optimal forecasts that attempt to exploit the same bias-variance trade-off that we have explored above. Indeed, in the special case in which $n = 1$, $\hat{\beta}^\star$ in (3.7) is identical to that from a Ridge Regression (Tikhonov estimator), which penalizes additional parameters with their $L2$ norm (Appendix B.4). The optimal estimator is also here equivalent to that in Eldar (2008), which following Stein's (1956) approach directly minimizes $trace\{mse\left(\hat{\beta} \mid x^t\right)\}$. Through this lens, our contribution is to employ the bias-variance trade-off *descriptively*, to argue that caution provides a rational response to the uncertainty that exists about the best use of limited information. We further show that the desire to exploit this trade-off does not rest on the presence of any cost function.[21]

We now turn to confronting our results with survey data on expectations.

4 Survey Data on Expectations

Uncertainty about the use of information can help account for several puzzling observations in survey data on expectations. In this section, we leverage our theoretical results to demonstrate that uncertainty about the optimal weight on information can account for the co-existence of biased-but-accurate expectations that over- and underreact to information. We end the section with two tests of our framework that explore the comparative statics of survey expectations.

[21] In Stein's (1956) analysis, which motivated the statistical learning literature, shrinkage towards any constant decreases the sum of the mean-squared errors of $p \geq 3$ estimates of the means of independent normally distributed variables. At a general level, the difference to our case, in which shrinkage towards zero is optimal, arises because the benefits of shrinkage towards some constant depend on the true parameters of the model. If an optimal "shrinkage constant" also has to be estimated from the data one returns to our case, analyzed above, in which all estimates are functions of the data, and where shrinkage towards zero is optimal.

4.1 Accuracy of Expectations

We start by exploring the accuracy of survey respondents' macroeconomic expectations, and its decomposition into a bias and variance component. We let $f_{it}y_{t+k}$ denote individual forecasts of y_{t+k} from a sample of respondents $i \in \{1, 2, \ldots, I\}$ at time t. An individual's forecast error is $y_{t+k} - f_{it}y_{t+k}$. We focus on output growth forecasts for two reasons. First, because expectations about future output play a central role in the economy as determinants of consumption, inflation, and asset prices. Second, because data on output forecasts are available for a longer time-span than forecasts of most other variables. Throughout, we rely on survey data on US output expectations from the *Survey of Professional Forecasters*.[22] Appendix C.1 studies the associated data on inflation expectations. We use real-time data to measure current realizations of output to capture the precise definition of the variable being forecasted.

A key implication of our framework (under the joint assumption that respondents observe the history of output realizations) is that respondents' expectations should be more accurate than those from ARIMA models estimated with (unbiased) maximum likelihood.[23] We consider a simple test of this prediction. Panel a in Figure 2 illustrates the average root mean-squared error (RRMSE) of one-quarter and four-quarter ahead output growth forecasts from the SPF relative to four different time-series models, two of which have been optimally-selected to best predict future output. A RRMSE ratio below one indicates that SPF forecasts are more accurate. We rely on the full-sample of observations to estimate the different models.

Although survey forecasts are biased, all time-series models fall short of survey forecasts at the one- and four-quarter horizon, with a pronounced difference at the one-quarter ahead horizon. This evidence suggests that respondents do better than simple linear models at forecasting output, consistent with Stark (2010), Faust and Wright (2013), and others.[24] Crucially, these empirical findings are in line with our theoretical results, in which agents down-weigh information to increase forecast accuracy. Indeed, Panel b in Figure 2 shows that the reason that survey forecasts outperform is precisely because they have lower error variance but are biased.[25] Survey respondents appear to exploit the basic trade-off between bias and

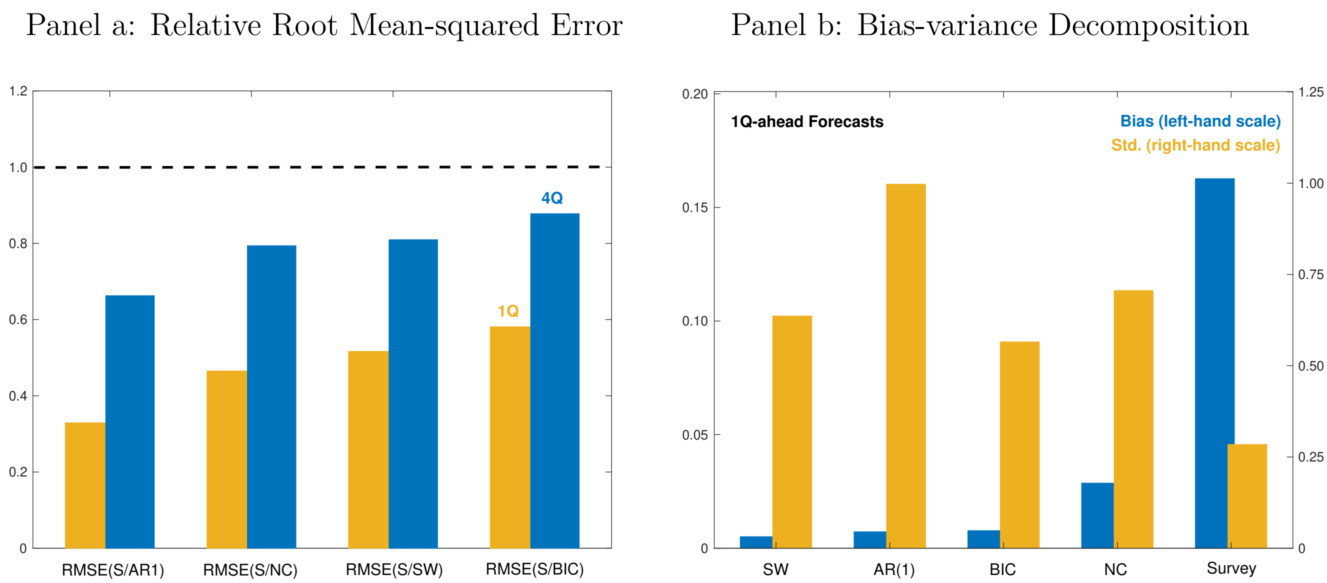
[22]The SPF is the oldest quarterly survey of individual US macroeconomic forecasts, dating back to 1968 (Croushore, 1993). Currently, the Federal Reserve Bank of Philadelphia administers the SPF and in each quarter surveys between 20-100 professional forecasters for their expectations over the next six-quarters.

[23]Because of the presence of lagged-dependent variables in ARIMA models the maximum likelihood estimator β^{ml} is slightly biased and inefficient. However, for the number of observations in the SPF sample, these effects are quantitatively very small, and one can in any case de-bias the maximum likelihood estimator to an arbitrary degree of accuracy using the approach in, for example, Tanaka (1984) or the Kendall-estimator in Orcutt and Winokur Jr (1969). The results in Figure 2 employ the former adjustment to de-bias the estimators.

[24]Updated values from Stark (2010) are available from the *Federal Reserve Bank of Philadelphia's* website: `https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/error-statistics`.

[25]Depending on whether an agent forecasts y_{t+k} or $-y_{t+k}$, as well as whether an agent is more or less

Figure 2: Forecaster Accuracy Relative to Time-series Models



Note: Panel a illustrates the average *relative root mean-squared error* of one-quarter and four-quarter ahead forecasts of year-over-year output growth from the US *Survey of Professional Forecasters* (S) relative to four time series models: *AR1* denotes forecasts from an AR(1) model, *NC* forecasts from a ("no-change") Random Walk, *SW* forecasts from an optimally-chosen time-varying parameter ARIMA model, using the method from Stock and Watson (2008), and BIC forecast from an ARIMA model chosen to minimize the BIC information criteria associated with one-quarter ahead forecasts. The use of the AIC information criteria selects the same model as BIC, which is why we exclude this alternative here. The sample period is 1970Q4:2020Q1. A RRMSE ratio below one indicates that SPF forecasts are more accurate. Panel b shows the decomposition of the model-implied root mean-squared errors of one-quarter ahead forecasts into a bias component (*Bias*) and the standard deviation of forecast errors (*Std.*) [see 3.2], and compares them to the survey data (*Survey*).

variance that is caused by uncertainty about the use of information to generate more accurate forecasts. Appendix C.1 shows similar patterns for respondents' inflation forecasts.

The evidence in Figure 2 and Appendix C.1 provides an important soundness check, which shows that our theory's main mechanism is consistent with basic moments of the survey data on professional forecasters. Our results further connect with the substantial body of evidence that documents biases in survey expectations (e.g., Elliott *et al.*, 2008). They are also in line with a strand of work, dating back to Muth (1961), showing that survey expectations often have limited variability. Our results above suggest a common explanation for these features: agents' responses to the uncertainty that exists about the optimal weight on information.

Finally, notice that any structural dynamic linear model (i.e., any linear DSGE or VAR

cautious to signals that, on average, decrease or increase her expectations, an agent's forecasts can be either positively or negatively biased (Appendix B.3). To avoid the resulting indeterminacy, we in this section focus on the "bias component" of (3.2), the absolute value of the bias, instead of its raw level. As the squared-value of the bias is all that enters our theoretical expressions (e.g., Proposition 3), this choice does not affect our subsequent analysis that attempts to match empirical moments with model-implied ones.

model) that is invertible has an ARIMA representation for output, and *vice versa* (Anderson and Moore, 1979; Fernández-Villaverde *et al.*, 2007). Such structural models condition output forecasts on more variables than the history of output observations. However, what is crucial for our purposes is that, because of their equivalent ARIMA representations, the forecasts from such models do not necessarily outperform those from simple time-series models (Mitchell *et al.*, 2019). In this sense, the class of models that we search over when selecting the most predictive models in Figure 2 is substantially richer than what may at first appear. This explains why our results are also consistent with those that show that SPF forecasts often outperform those from linear DSGE or VAR-based models (e.g., Bennett and Owyang, 2022; Faust and Wright, 2013). We further note that the above estimates are also *prima facie* inconsistent with models in which agents derive forecasts from misspecified ARIMA models (e.g., Fuster *et al.*, 2010), or adjust their use of information in a solely behavioral manner.

Combined, the results in Figure 2 demonstrate that uncertainty about the optimal weight on information can *qualitatively* account for the bias-accuracy trade-off visible in survey data on expectations. Figure 3 makes a first pass at a *quantitative* description.

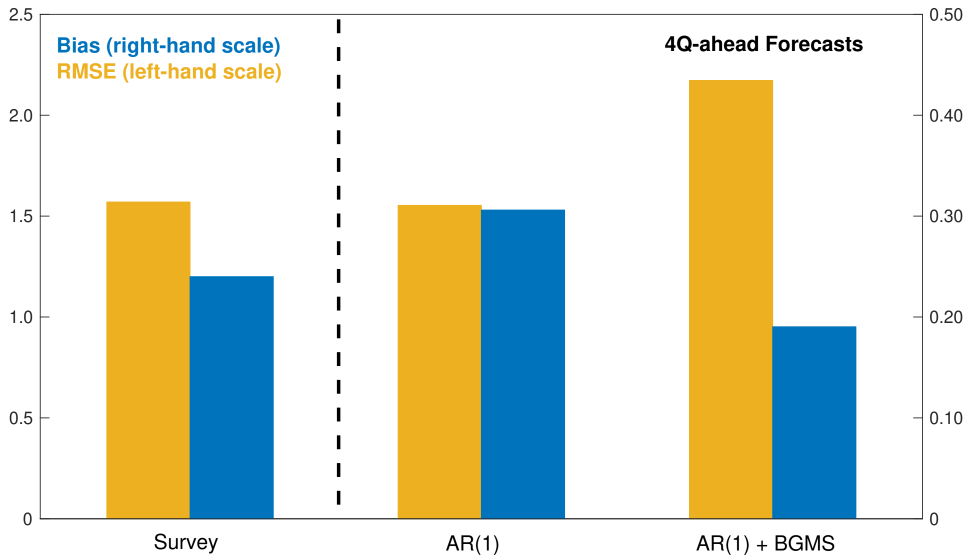
The figure shows the model-implied root mean-squared error and bias component of four-quarter ahead expectations of output growth, assuming that output growth follows an AR(1) process with parameters equal to those estimated in Bordalo *et al.* (2020). The signals that respondents use to forecast future output growth y_{t+k} are thus equal to a constant and current output, $x_t = [\ 1 \quad y_t\ ]'$. The figure further assumes that respondents base their forecasts on $T = 14$ observations of output,[26] consistent with the median tenure of respondents in the SPF who have participated for at least two quarters.[27] We will later use this parametrization to also study respondents' over- and underreactions to new information.

Figure 3 compares the average model-implied estimates to the survey data. Although model parameters are not targeted to match the accuracy of survey forecasts, the model captures well the root mean-squared error of forecasts in the survey data, consistent with the mechanisms discussed in Section 3. The model also accurately matches the overall level of bias in errors, although it entails somewhat too skewed forecasts. Overall, we thus find that the

[26] Because of the presence of lagged-dependent variables, the maximum likelihood/least-squares estimator is slightly inefficient and biased with $T = 14$ observations. We use the Kendall-adjustment to de-bias the estimator and increase its efficiency (Orcutt and Winokur Jr, 1969). This provides us with (an approximately) unbiased minimum-variance transformation of the maximum likelihood estimator, which we use as the replacement for β^{ls} in Proposition 3. Tanaka (1984) shows that the Kendall-adjustment provides an accurate adjustment for cases with $T > 10$ and moderate-to-high persistence. Appendix C.2 further describes a small change in the timing convention from Section 2 necessary to deal with the AR(1) case.

[27] Because the SPF includes a larger number of forecasters that are only in the survey for one round, we only consider forecasters that are in the survey for at least two consecutive quarters. The raw median, including respondents who are only in the survey for one quarter, is $T = 8$. Furthermore, because caution choices $m^\star$ increase with time, we throughout adopt the conservative assumption that $m^\star$ is fixed at its maximal value, which corresponds to the full sample of observations.

Figure 3: Optimal Expectations based on an AR(1) Process



Note: The left-two columns show the root mean-squared error and bias component (equation 3.2) of four-quarter ahead forecasts of year-over-year output growth from the *SPF* (Survey). The sample period is 1970Q4:2020Q1. The middle-two columns (AR(1)) demonstrate the average model-implied root mean-squared error and bias component of four-quarter ahead optimal expectations (Proposition 3), assuming that y_t follows an AR(1) with persistence $\rho = 0.85$, standard deviation of innovations $\sigma_\eta = 0.85$, and mean $\mu = 2.40$. We further assume that respondents base their forecasts on $T = 14$ observations of output, consistent with the median tenure of respondents in the SPF who have participated for at least two quarters. The right-two columns (AR(1)+BGMS) show the model-implied root mean-squared error and bias component, assuming that respondents observe a noisy signal of output growth, $x_{it} = y_t + \epsilon_{it}$, $\epsilon_{it} \sim \mathcal{N}\left(0, \sigma_{x2}^2\right)$, where $\sigma_{x2} = 1.20$, in each period instead of a noiseless signal. Subsection 4.2 describes the calibration of σ_{x2}.

model-implied errors have comparable quantitative properties to their empirical counterparts, despite these not being part of the parametrization. We conclude, as a result, that our simple framework can quantitively match the bias-accuracy trade-off in the survey data.

We next turn to two tests of our framework that explore respondents' reactions to new information, in addition to their bias and accuracy.

4.2 Over- and Underreactions of Expectations

We leverage our results on the predictability of forecast errors to demonstrate that uncertainty about the optimal weight on information is also consistent with respondents' over- and underreactions to information (Coibion and Gorodnichenko, 2015, Bordalo *et al.*, 2020).

A well-known implication of full-information and rational expectations (with mean-squared error preferences) is that individual forecast errors should be unpredictable. Under rational expectations with a known optimal use of information, no variable observable at time t should correlate with $y_{t+k} - f_{it}y_{t+k}$. The left-hand panel of Table I tests this prediction.

The table reports estimates from a regression of individual errors onto respondents' *average* and *individual* revisions ($I^{-1}\sum_i [f_{it}y_{t+k} - f_{it-1}y_{t+k}]$ and $f_{it}y_{t+k} - f_{it-1}y_{t+k}$), respectively. Under full-information and rational expectations, respondents' errors should be uncorrelated with either, as both average and individual revisions are by assumption observable at the time of the forecast. If respondents instead overreact to average and individual information received between two periods, displaying overoptimism in response to positive news about output, then future errors should, by contrast, be negatively correlated with both. The table also includes our earlier measure of bias from a regression of errors onto a constant term.

Consistent with results in Coibion and Gorodnichenko (2015), Bordalo *et al.* (2020), and Broer and Kohlhas (2022),[28] columns (2) and (3) document *simultaneous under- and overreactions*. Expectations *underreact* to the average information observed between two periods, consistent with a positive correlation between errors and average revisions. Yet, expectations conversely *overreact* at the individual-level, as shown by a negative relationship between errors and individual revisions. Taken together, the estimates in Table I provide strong evidence of deviations from the full-information rational expectations benchmark. Indeed, at first pass, the estimates seem to contradict the rational use of information. However, as we show below, if agents are uncertain about the optimal weight on information, this pattern of over- and underreactions can also arise as an optimal outcome.

To illustrate the conditions under which this occurs, we borrow the baseline setup from Bordalo *et al.* (2020). A continuum of agents $i \in [0, 1]$ forecast output y_{t+k}, which evolves in accordance with an AR(1) with mean μ and persistence $\rho \in (0, 1)$. However, unlike in the previous subsection, at the start of each period, each agent observes her own noisy signal of current output $x_{i2t} = y_t + \epsilon_{it}$, where $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2_{x2})$, in addition to the conditional prior expectation $x_{i1t} \equiv \mathbb{E}\left[y_t \mid x^{t-1}_{i2t}; \beta\right]$ with mean-squared error σ^2_{x1}. The AR(1)-model from the previous subsection can be seen as the special case in which $\sigma_{x2} \to 0$. Appendix C.2 shows that the noisy-information setup also falls into our framework from Section 2. In particular, we have $x_{it} = \begin{bmatrix} 1 & x_{i1t} & x_{i2t} \end{bmatrix}'$ with $\beta_1, \beta_2 > 0$ and σ^2 that is a function of σ^2_{x1} and σ^2_{x2}.

Because of uncertainty about the optimal weight on information, agents' expectations are generically biased (as in column (1) in Table I). Furthermore, because of the presence of noisy, individual-specific information, agents' expectations are also consistent with the documented *underreaction* to average revisions (column (2) in Table I). The noisiness of individual infor-

[28] Coibion and Gorodnichenko (2015) estimate regressions of *average* forecast errors on *average* forecast revisions. However, as pointed out by Kohlhas and Walther (2021), such regressions are asymptotically equivalent to regressions of *individual* forecast errors on *average* forecast revisions. Furthermore, in finite samples, the individual-level version has several statistical advantages, which is why we choose to adopt this specification. See also the discussions in Kohlhas and Walther (2021) and Angeletos *et al.* (2021) about the influence of outliers on the estimates presented in Table I.

Table I: Over- and Underreactions of Optimal Expectations and Data

	Data Moments			*Model Moments*		
	(1)	(2)	(3)	(1)	(2)	(3)
Constant	0.24***	–	–	0.19	–	–
	(0.03)			(–)		
Average Revision	–	0.68***	–	–	0.68	–
		(0.19)			(–)	
Individual Revision	–	–	-0.19***	–	–	-0.33
			(0.06)			(–)
Observations	7,104	7,065	5,469	.	.	.

Note: The left-hand panel ("Data Moments") shows estimates of a regression of $y_{t+k} - f_{it}y_{t+k}$ on respondents' average- and individual forecast revisions ($I^{-1}\sum_i [f_{it}y_{t+k} - f_{it-1}y_{t+k}]$ and $f_{it}y_{t+k} - f_{it-1}y_{t+k}$), respectively. Columns (2) and (3) include individual (respondent) fixed effects and $k = 4$. Double-clustered robust standard errors in parentheses. Sample: 1970:Q4–2020:Q1. The right-hand panel shows average model-implied moments. The accuracy of model forecasts and their bias can be seen in the right-two columns of Figure 3.

mation, all else equal, causes agents to underreact to the average information observed in the population.[29] However, what is surprising is that the caution choices associated with uncertainty about the optimal weight on information can also help reconcile our model with the documented *overreaction* to individual revisions (column (3) in Table I).

Proposition 7. *Let $\hat{\beta}^\star$ be fixed at its mean value $\left(\hat{\beta}^\star = m^\star_{[k]}\beta\right)$ with $m^\star_{[k+1]} \in \left(m^\star_{[k]}, \rho^{-2}m^\star_{[k]}\right)$, where $m^\star_{[k]}$ denotes the caution choice for the kth-horizon forecast, and consider the coefficient γ from a regression of $y_{t+k} - f_{it}y_{t+k}$ on $f_{it}y_{t+k} - f_{it-1}y_{t+k}$.*

1. *Individual errors underreact to individual revisions ($\gamma > 0$) when the true weight on the signal x_{i2t} is large ($\beta_2 \to \rho^k$) relative to the prior x_{i1t} ($\beta_1 \to 0$).*
2. *Individual errors overreact to individual revisions ($\gamma < 0$) when the true weight on the signal x_{i2t} is small ($\beta_2 \to 0$) relative to the prior x_{i1t} $\left(\beta_1 \to \rho^k\right)$.*

Proposition 7 shows that whether agents under- or overreact to individual revisions depends on the relative weight on prior versus new information. On the one hand, *caution* to *new information* $\hat{\beta}^\star_2 = m^\star_{[k]}\beta_2$ with $m^\star_{[k]} \in (0, 1)$, all else equal, leads agents to *underreact*, as agents

[29]Consider the case with a known optimal use of x_{it}, in which $f_{it}y_{t+1} = \mathbb{E}\left[y_{t+1} \mid x^t_{i2}; \beta\right]$. As shown by Coibion and Gorodnichenko (2015), average errors and their individual counterparts are, in this case, positively correlated with average revisions. Expectations rationally underreact to the average new information observed in the population (equal to $\int_0^1 x_{i2t}di = y_t$), due to the cancelling of noise in individual signals. This force, all else equal, also leads to a positive correlation between individual errors and average revisions in our case, where agents are uncertain about the optimal use of information.

optimally down-weigh their responses to the noisy signal x_{i2t}. This, all else equal, leads to a positive correlation between individual errors and revisions, and amplifies any underreactions to average revisions. On the other hand, however, *caution to prior expectations* $\hat{\beta}_1^\star = m_{[k]}^\star \beta_1$ with $m_{[k]}^\star \in (0, 1)$ leads agents' expectations to *overreact.*

Intuitively, when agents exhibit caution towards prior expectations, their forecasts place only a small weight on their moderating force. As a result, when new information x_{i2t} is high agents, on average, tend to revise up their expectations by more than an agent who knows the optimal use of information. This leads to a seeming overreaction to new information, which manifests itself in a negative correlation between future errors and current revisions. In this sense, overeactions arise because an overreaction to new information can be interpreted as an underreaction (i.e. an extra down-weighing) of prior expectations relative to new information.

Proposition 7 shows that the horizon profile of caution factors ($m_{[k]}$ vs. $m_{[k+1]}$) combines with the effects stemming from the relative weight on prior and new information to determine whether agents under- or overreact to individual revisions. In particular, depending on the relative weight on prior and new information, more or less shrinkage of current versus prior expectations can advance over- or underreactions. More emphasis on prior expectations, all else equal, furthers overreaction to individual revisions.

Lastly, we note that the insights from Proposition 7 are in line with recent work that explores the comparative statics of individual-level overreactions. As in Kwon and Tang (2022), overreactions here arise when new information is inaccurate and uncertainty is high, such as after novel or extreme circumstances. Proposition 7, moreover, makes the new prediction that overreactions are more likely when learning has been limited and caution factors differ more from one another across horizons.

Although the model framework considered in this section is simple, the right-hand panel of Table I explores the capacity of our framework to also *quantitatively* match the survey evidence. To do so, we set the parameters that govern output growth y_t equal to those in Section 4.1. We calibrate the noise σ_{x2} in x_{i2t} to target the average correlation between individual errors and average revisions. The noisiness of new information, all else equal, also controls the relative weight on prior x_{i1t} versus new x_{i2t} information in agents' expectations.

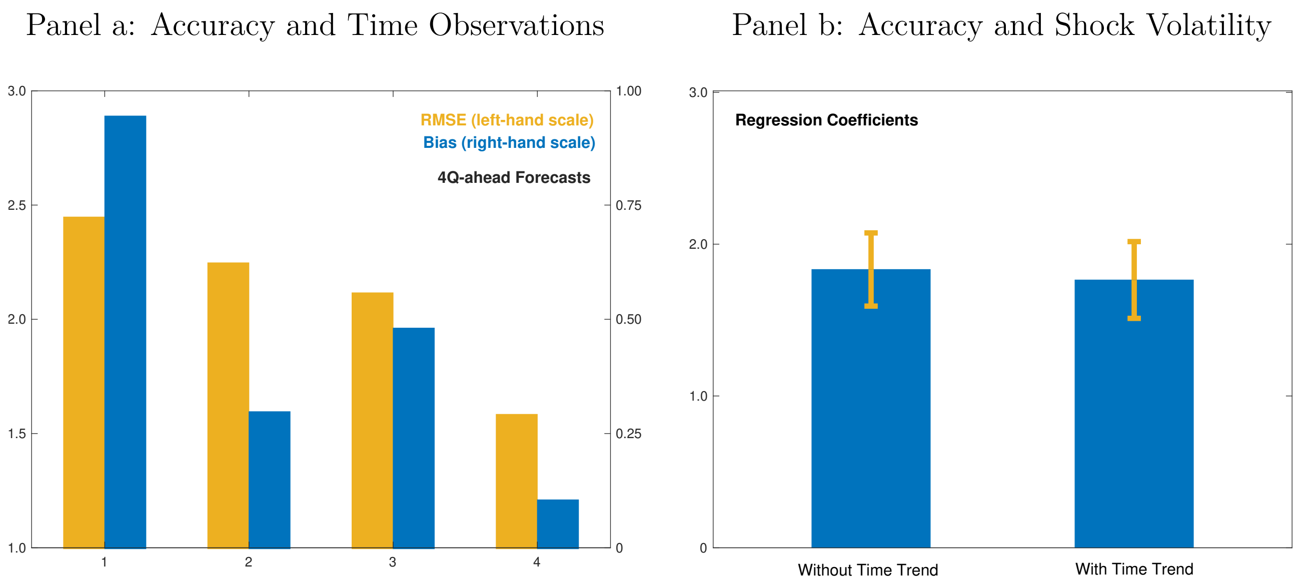
The right-hand panel of Table I shows that our model is able to capture all three moments well. We estimate a smaller bias than without the presence of noise in the signal of current output, closer to that in the survey data (Figure 3). We also estimate overreactions to individual revisions that are somewhat larger than in the data. However, as discussed in Bordalo *et al.* (2020) and Broer and Kohlhas (2022), among others, the uncertainty around these estimates is large. Finally, we note that our model, calibrated in this manner, still captures the estimated accuracy of one-year ahead output growth forecasts (Figure 3). We conclude that our simple framework can simultaneously account for respondents' over- and underreactions at the same time as the accuracy and bias of expectations. Appendix C.1 shows that our model can also capture similar salient features of respondents' inflation forecasts.

In summary, our model is able to account for stylized facts about the accuracy, bias, and responsiveness of survey expectations. This occurs when prior expectations are important for individual beliefs, so that the down-weighing of prior expectations dominates that which occurs towards new information. These results chime in with a large psychological literature that has documented "prior or base-rate neglect" in experimental settings (e.g., Phillips and Edwards, 1966 and Kahneman and Tversky, 1973). As in, for example, Benjamin (2019), agents' calibrated expectations in Table I systematically down-weigh prior expectations relative to new information, which generates overreactions similar to those in the survey data. Finally, our results cast new light on previous explanations for individual-level overreactions, in which respondents overreact for behavioral reasons (e.g., Bordalo *et al.*, 2020). Our results demonstrate that what is important for matching the survey data is not whether the weight on new information is above or below the rational value assuming a known optimal weight on information. But rather, what matters is what the relative weight is compared to that on prior expectations. Overreactions can arise solely from the down-weighing of prior expectations.

4.3 Comparative Statics of Expectations

We close our analysis of the survey data by briefly exploring the comparative statics of the accuracy and bias of survey expectations. To do so, we exploit the SPF survey to show that respondents who are in the survey for longer, a proxy for the number of observations used in their forecasts, make more accurate predictions with a smaller bias (Panel a in Figure 4). We also show that the larger the variance of shocks to output growth is, proxied by an increase in the conditional variance of output growth from a GARCH(1,1) model, the worse respondents forecasts become (Panel b in Figure 4). Crucially, this deterioration occurs at a rate greater than one-for-one, consistent with an increase in the second term of (3.1) that measures accuracy losses due to parameter uncertainty. Although these exercises push the limits of what is feasible within the SPF survey, the overall picture nevertheless still conforms with agents' responses to uncertainty about the use of information: more observations lead to more accurate (i.e. lower mean-squared error) and less biased forecasts, while the converse is true for periods with a larger volatility of shocks. We interpret this supplementary evidence as further validation of our framework, which implies that our theory is also consistent with key dimensions of the comparative statics of survey expectations.

Figure 4: Comparative Statics of Forecast Accuracy



Note: Panel a illustrates the root mean-squared error and bias component of four-quarter ahead forecasts of year-over-year output growth from the US Survey of Professional Forecasters as a function of the number of quarters an individual respondent has been in the survey: "1" (the bottom 25 percent of the distribution of quarters); "2" (between the 25th and the 75th percentile); "3" (between the 75th and 95th percentile); "4" (the top 5% percent of the distribution of quarters). The median number of quarters is $T = 14$. Panel b shows the coefficient estimate from a linear regression of the h−quarter rolling mean-squared error of one-year ahead forecasts on the estimated shock volatility from a GARCH (1,1) estimated on output growth. We set h = 15 and assume an AR(1) process for the mean of output growth. Estimates are shown with and without a linear time-trend. Whisker-intervals correspond to 95 percent robust confidence bounds. Sample: 1970Q4-2020Q1. Appendix Table C.1 provides further details on the estimation.

4.4 Summary and Discussion

In this section, we have documented that uncertainty about the optimal use of information can help account for salient features of survey data on professional forecasters' macroeconomic expectations. In particular, we have demonstrated that uncertainty about the optimal weight on information can simultaneously match the accuracy, bias, and over- and underreactions of survey expectations. Our results contrast with several recent models of expectation formation, in which expectations either (i) over- and underreact, but are unbiased and less accurate (e.g.; "diagnostic expectations"; Bordalo *et al.*, 2020); or (ii) are biased, but do not simultaneously over- and underreact to information (e.g., "strategic models of forecasters"; Marinovic *et al.*, 2013).[30] Although the list of models which combines elements from either class is not exhaustive, we are unaware of a pre-existing model that can explain our results.

[30]See also Scharfstein and Stein (1990), Bordalo *et al.* (2017), Kohlhas and Walther (2021), Angeletos *et al.* (2021), Farmer *et al.* (2022), Da Silveira *et al.* (2022), Gemmi and Valchev (2022), and Sung (2022) for both behavioral or rational models of either type.

Clearly, in any given application, the extent to which our framework's expectations will differ from those which assume a known optimal weight on information will depend on the number of time-series observations and signals observed, the noisiness and structure of the data, as well as the cross- and own information structure of signals (Section 3.4). As shown above, several of these elements can be taken directly from the data.

One unexplored advantage of our framework is its capacity to scale, and to be applied to more complex environments with many signals. Indeed, Proposition 3 to 5 can readily be applied to any linear-normal prediction problem with an arbitrary number of signals with little effort. Appendix B.2 shows how the basic insights from Section 3 also extend to non-normal and non-linear settings. This in principle allows our theory to match a broader set of survey facts. All else equal, more complex environments with many different signals will increase agents' implied caution choice ($m^\star \downarrow$) and lead to larger differences between agents' expectations and those from standard benchmarks with a known optimal weight on information.

Finally, one limitation of the above applications is that they do not illustrate how cross-information alters agents' expectations, nor how we can straightforwardly allow caution choices to vary across signals. Appendix B.3 and D.4 demonstrate this potential.

5 A Consumption-Saving Problem

In this section, we illustrate the consequences of uncertainty about the optimal use of information in a standard consumption-savings problem. We choose this application because consumption and saving choices lie at the bedrock of workhorse macroeconomic models. We analyze an environment in which an agent with quadratic preferences has to estimate the best use of knowledge of her own productivity to predict future income. We show that the agent's problem maps into a special case of our framework. We document that, for standard parameters, the agent optimally chooses an upward-sloping consumption profile and muted, persistent responses to income shocks. Finally, we show that increased riskiness of future income results in a sizable decline in current consumption, despite linear decision rules, and map the decline into an *as-if* estimate of risk aversion. Appendix D.5 applies our ideas to a model of monetary policy in the spirit of Maćkowiak and Wiederholt (2009).

5.1 Model Setup

We consider the behavior of a household in a multi-period consumption-savings problem with stochastic income and time-separable, quadratic utility. In an initial period $t = 0$, the agent chooses *ex-ante* how to best construct expectations $f_t[\cdot]$ at later dates $t > 0$. In each subsequent period, $t = 1, 2, \ldots, T$, the agent then chooses consumption $\{c_t\}$ to maximize her own expectation of the realized, discounted utility from consumption,

$$\sum_{t=1}^{T} \delta^{t-1} u\left(c_t\right), \tag{5.1}$$

subject to the budget constraint

$$\sum_{t=1}^{T} R^{-t}\left(c_t - y_t\right) = a_1, \quad a_1 \geq 0, \tag{5.2}$$

where $\delta \in (0,1)$, $u(\cdot)$ is quadratic with $u' > 0$ and $u'' < 0$ over the relevant range for $c \geq 0$, and initial wealth $a_1 \geq 0$. We also, for simplicity, assume that $\delta R = 1$. The only uncertainty the agent faces at time t is over future income y_{t+k}, which depends upon her own current and future productivity z_{t+k-1}:

$$y_{t+k} = \beta_1 z_{t+k-1} + \eta_{t+k}, \quad \eta_{t+k} \sim \mathcal{N}\left(0, \sigma^2\right), \tag{5.3}$$

where $\beta_1 > 0$, $\mathbb{E}\left[\eta_{t+k}\eta_{s+k}\right] = 0$ for all $s \neq t$, and

$$z_{t+k} = \lambda z_{t+k-1} + u_{t+k}, \quad u_{t+k} \sim \mathcal{N}\left(0, \sigma_u^2\right), \tag{5.4}$$

with $\lambda \in (0,1)$, $\mathbb{E}\left[u_{t+k}u_{s+k}\right] = 0$ for all $s \neq t$, $\mathbb{E}\left[u_t \eta_s\right] = 0$ for all s, t, and $z_0 > 0$. The agent's information set at $t \geq 1$ is $\Omega_t = \{y^t, z^t; \delta, R, \lambda, \sigma_e, \sigma_u\}$.[31] The evolution of income in (5.3) and (5.4) is a special case of our framework in (2.2) with $n = 1$ and $x_{1t} = z_t$: the agent uses current realizations of productivity x_{1t} to best predict next-period's income y_{t+1}. We note that the agent does not know the mapping β_1 between x_{1t} and y_{t+1} in (5.3). Instead, the β_1-coefficient must once again be estimated from the history of observations.

5.2 Consumption and Expectation Choices

We start by characterizing the agent's consumption choice at $t \geq 1$, assuming her expectations satisfy the Law of Iterated Expectations. Appendix D.1 and D.2 demonstrate that we can solve the agent's problem using standard steps. Assuming an interior solution shows that the optimal consumption choice depends on the agent's expectation of life-time resources:

$$c_t = \theta_t \left(Ra_t + y_t + \sum_{k=1}^{T-t} R^{-k} f_t\left[y_{t+k}\right]\right) = \theta_t \left(Ra_t + y_t + \sum_{k=1}^{T-t} \lambda^{k-1} R^{-k} f_t\left[y_{t+1}\right]\right), \tag{5.5}$$

[31] We, furthermore, assume that the agent observes (x_0, y_1) and (x_{-1}, y_0) in the initial period, so that the agent can construct an estimate of β_1 in period $t = 1$.

where $\theta_t \equiv \frac{1-R^{-1}}{1-R^{-(T-t+1)}}$ and the second equality uses that $f_t y_{t+k} = \lambda^{k-1} f_t y_{t+1}$.[32]

We can use equation (5.5) to solve the agent's expectation-formation problem at $t = 0$. Let $c_t^\star$ denote the perfect-foresight value of c_t in (5.5). Appendix D.2 shows that the agent's consumption-savings problem in (5.1) and (5.2) can then equivalently be stated as choosing consumption $\{c_t\}$ to maximize the agent's expectation of $\sum_{t=1}^{T} \delta^{t-1} \left[-\frac{1}{2}\left(c_t - c_t^\star\right)\right]$. Inserting the condition for $c_t^\star$ into this expression and expanding terms shows that the agent will choose her expectation $f_t y_{t+1}$ at $t \geq 1$ to maximize

$$-\frac{1}{2}\mathbb{E}\left[y_{t+1} - f_t y_{t+1}\right]^2. \tag{5.6}$$

We conclude the agent's expectation-formation problem is nested within our framework. As a result, $f_t y_{t+1} = f_t^\star y_{t+1}$, where $f_t^\star y_{t+1}$in (3.7) abides by the Law of Iterated Expectations.

5.3 Consumption Implications

We leverage the solution to our consumption-saving problem to illustrate three wider implications of uncertainty about the optimal use of information. First, we show that uncertainty about the optimal weight on information leads to upward-sloping consumption profiles and predictable consumption changes. Second, we show that decreased predictability of future income results in a sizable decline in current consumption, despite a linear decision rule, and map the decline into an *as-if* estimate of risk aversion. Both findings contrast with the case in which the agent knows the optimal weight on information. Finally, we demonstrate that uncertainty about the best use of information leads to dampened, persistent responses to income shocks, and relate our findings to the evidence on the effects of stimulus checks during the Great Recession.

5.3.1 Consumption Levels and Changes

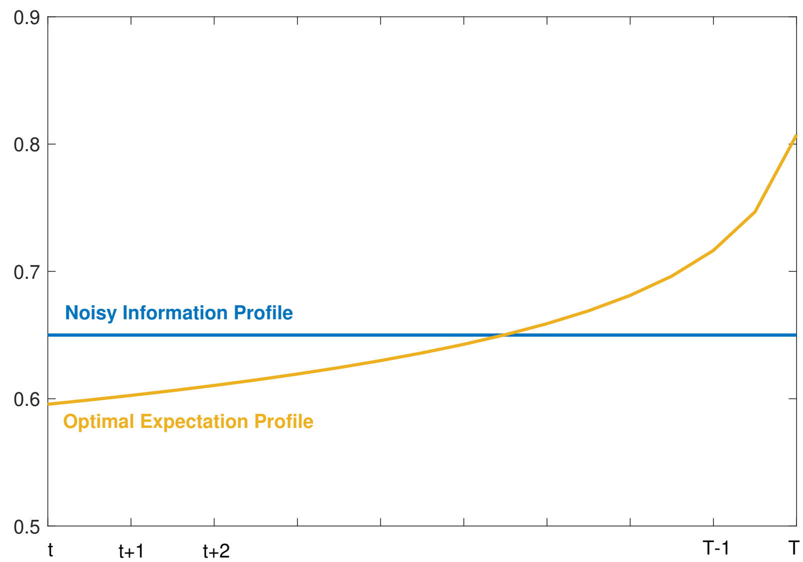
We compare the average level of consumption across time in our model with that which arises in a model in which the agent knows the optimal weight on information but is unsure of the breakdown of income into a persistent and transitory component. In particular, in this noisy information case, the agent knows β_1 in (5.3) but must instead infer movements in productivity x_{1t} from observations of income y_t.[33] The rest of the model is unchanged.

Corollary 1 summarizes our results, using our earlier findings in Proposition 3 and 5 combined with equation (5.5). Figure 5 illustrates the average life-cycle consumption profile for

[32]The agent's assets at time $t \geq 2$ are equal to $a_t = Ra_{t-1} + y_{t-1} - c_{t-1}$.

[33]Furthermore, in this noisy information case, we assume that the agent receives an additional signal $\tilde{y}_0$ at the start of time that initializes her beliefs at the steady state of the Kalman filter recursion associated with her signal extraction problem. This is in line with the related literature (see, e.g., Maćkowiak *et al.*, 2021).

Figure 5: Average Life-Cycle Consumption Profiles



Note: The figure depicts the average life-cycle consumption profile from (5.5) (Optimal Expectations Profile) and compares it with the noisy information counterpart described in the main text (Noisy Information Profile). The figure considers an agent who has observed income and her own productivity for 13 periods ($t = 13$) and will live for another 20 periods ($T = 33$). The income process y_t is calibrated to match the variance and autocorrelation estimated for output in Section 4 ($\beta = 0.65$, $\lambda = 0.96$, $\sigma_e = \sigma_u = 1/2$), we set $R = 1$, and we throughout fix $x_{1t} = 1$.

an agent, using the calibrated process for output from Section 3.6.

Corollary 1. *Suppose productivity is held fixed $x_{1t} = x > 0$. The agent then optimally chooses an upward-sloping consumption profile, on average. This is despite income being constant, on average, and contrasts with the constant choice in the noisy-information case. Furthermore, consumption choices in (5.5) exhibit myopia and changes are predictable based on income.*

There are two forces that push the agent to choose an upward-sloping consumption profile in Corollary 1, both of which are caused by the agent's caution choice $m^\star \in (0, 1)$. First, at the start of the agent's life, where t is relatively small and the agent has few observations of her income, the agent optimally chooses to shrink future income expectation relatively more. The implied caution choice $m^\star$ is increasing in t (Proposition 5). This, all else equal, causes the agent to choose lower consumption at the start of her life and save more, which can be viewed as a type of *precautionary savings* that is distinct from that caused by any prudence of the utility specification (Kimball, 1990). Second, because the agent chooses to save more initially, she also has higher consumption later on in life.[34]

[34]Suppose we were to introduce a probability of death $\tilde{\beta}_2 \in (0, 1)$ in each period, unknown to the agent,

Combined, these two properties ensure that consumption choices are on average increasing (Figure 5), consistent with the empirical evidence (e.g., Gourinchas and Parker, 2002). Notice that we do not require income to be related to experience, financial frictions, or any other feature for consumption to increase with age. By contrast, in the noisy-information case, where the agent is unsure of the breakdown of income into a persistent and transitory component, her optimal (conditional) expectation $f_t\left[y_{t+k}\right] = \mathbb{E}\left[y_{t+k} \mid y^t; \beta_1\right]$ averages to mean income $\mathbb{E}\left[y_{t+k}\right]$, and the agent, on average, chooses the same consumption profile as under full-information and rational expectations (FIRE).

Turning to consumption *changes* rather than *levels*, another noticeable feature of (5.5) is that consumption changes are predictable. The agent's consumption choices satisfy the Euler equation $f_t\left[c_{t+1}\right] - c_t = 0$ (Appendix D.2). But since the agent's expectations differ from the conditional expectation assuming knowledge of β_1, consumption changes are not martingales. Instead, in line with the evidence on violations of the "Hall (1978) martingale conjecture", consumption changes are predictable based on current income (Section 3.6). This again contrasts with the noisy and full-information cases.

Finally, we note that the consumption choices described in Corollary 1 entail a form of myopia, which resembles that in Gabaix (2017) and Angeletos and Huo (2021). Expectations about the future matter for consumption in (5.5) through the term $\sum_{k=1}^{T-t} R^{-k} f_t\left[y_{t+k}\right]$. The FIRE value of this expression is $\bar{y}_{t,FIRE} \equiv \sum_{k=1}^{T-t} \lambda^{k-1} R^{-k} \beta_1 x_t$, which is strictly greater than the expected value of the agent's optimal expectation $\mathbb{E}\left[\sum_{t=1}^{T-t} R^{-k} f_t\left[y_{t+k}\right] \mid x_1^t\right] = m^\star \bar{y}_{t,FIRE}$ as $m^\star \in (0,1)$. In this sense, consumption choices exhibit myopia: expectations about future income matter, on average, less for today's consumption than in the FIRE case. We next turn to the implications of our framework for the economy's responses to shocks.

5.3.2 Consumption Responses to Shocks

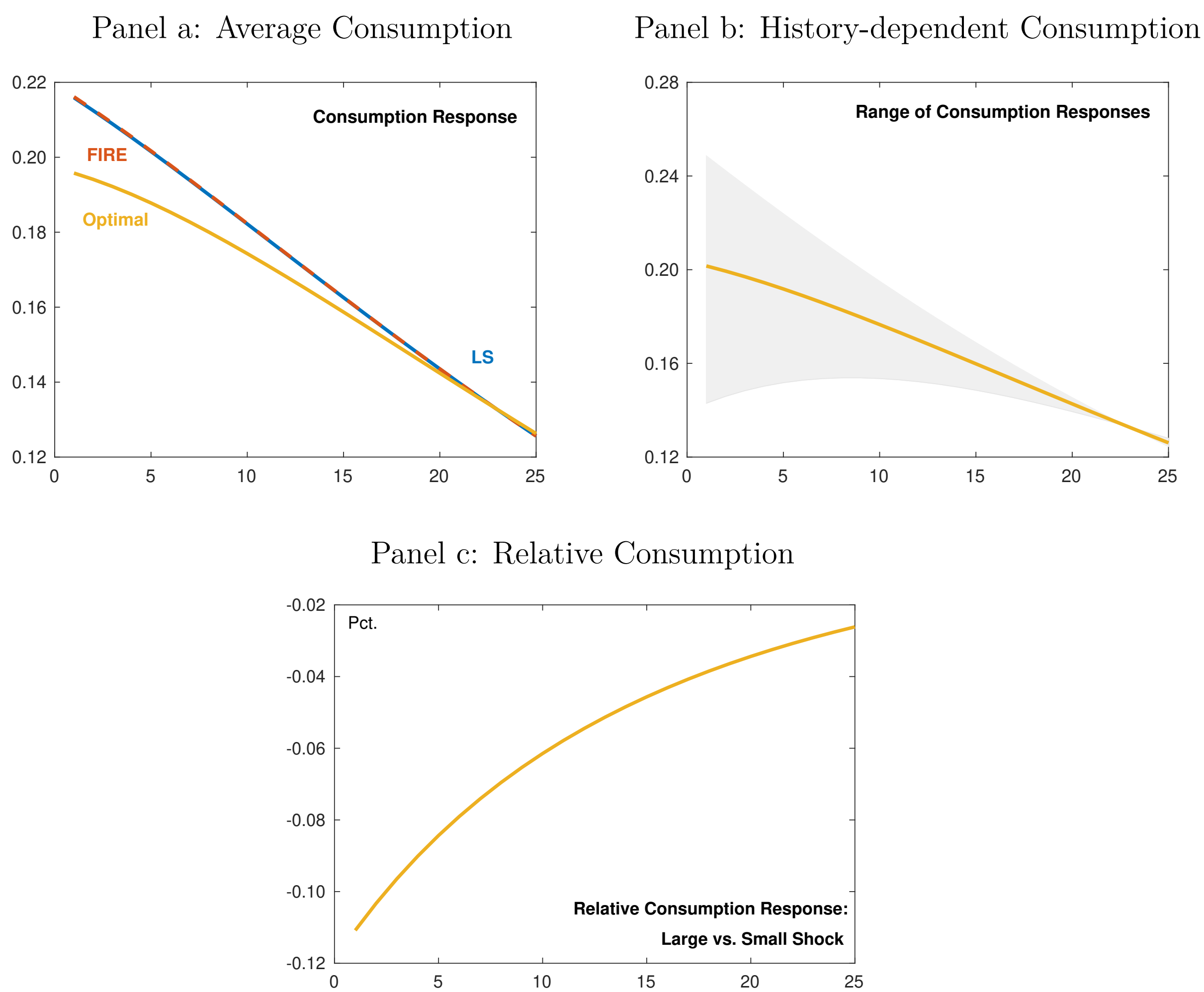
Recall from Proposition 3 and 5 that (i) uncertainty about the optimal weight on information causes a dampened response to signal realizations, and that (ii) the magnitude of this decrease depends on the realized volatility of signals. Corollary 2 and Figure 6 illustrate the consequences of these mechanisms for our economy's response to a productivity shock.

Corollary 2. *The consumption response c_{t+k} to a productivity shock u_t is initially smaller, on average, than under knowledge about the optimal weight on productivity x_{1t}, and is more persistent. Furthermore, consumption responses are larger after large movements in x_{1t}.*

---

instead of a fixed terminal end-date T. As with the productivity parameter β_1, the agent would optimally choose to shrink this coefficient when forecasting future realizations. We conjecture that this, in turn, would imply that the agent would, on average, die with strictly positive wealth.

Figure 6: Consumption Responses to Income Shocks



Note: The figure illustrates the impulse response of consumption in (5.5) to increases in productivity. Panel a shows the average impulse response to a one standard deviation increase when the agent's expectations follows (3.7) and compares it to the FIRE and least-squares/ML case. Panel b depicts the 17th and 83rd percentile of realizations of the impulse response function across different simulations for productivity and income. Finally, Panel c illustrates the relative impulse response of consumption: we compare the average response of an economy for $\tau \geq t$ that experiences a one and two standard deviation increase in the transitory component η_t of income in period t, respectively, and fix $x_{1t} = 1$ throughout. A negative value shows a larger response in the two-standard deviation case. All parameters are equal to those in Figure 5.

That consumption responses are initially muted is a consequence of the decreased weight the agent, on average, places on productivity x_{1t} in her expectation of future income $f_t y_{t+k}$. After a positive productivity shock, the resulting smaller increase in income expectations causes the agent to increase consumption by less—this is relative to an agent who is certain of β_1. What is interesting is that the consumption response is also more persistent. This is because an agent who is uncertain of β_1 accumulates more savings after the shock occurs. As a result, productivity shocks have more protracted effects on consumption (Figure 6 Panel a).

The results in Corollary 2 and Panel a in Figure 6 closely resemble those in the noisy-information case, where the agent uses realizations of income y_t to estimate underlying productivity x_{1t}, the latent factor. Similar to this framework, consumption responses are muted and more persistent because the agent does not fully update her expectations about future income in response to shocks. However, Corollary 2 and Figure 6 also demonstrate important differences to models in which noisy information is the underlying friction.

First, unlike in, for example, Sims (2003), the consumption effects of productivity shocks are history-dependent. The precise realization of x_1^t matters for the agent's optimal weight $\hat{\beta}_1^\star$, and hence for how the agent chooses to down-weigh her responses (Proposition 3). Panel b in Figure 6 illustrates the range of outcomes that can arise after a one-standard deviation increase in productivity. Notice that the range of outcomes is substantial, depending on the precise realization of past productivity shocks. All else equal, economies that have experienced more volatile fundamentals x_1^t will show larger, less persistent increases in consumption. This is because agents in these economies are, on average, more responsive to information, and hence adjust their consumption by more.

Second, this history-dependence can also help explain facets of the data that seem at odds with simple noisy information models. For example, Johnson *et al.* (2006) and Parker *et al.* (2013) document that the economic stimulus payments in 2008 had a smaller "per-dollar" impact on U.S. non-durable consumption than the smaller 2001 stimulus package. This is inconsistent with the outlined noisy-information framework, where the "scaled" effects of a transitory shock η_t are constant in size. Due to the history-dependent use of information, our model can, however, speak to this evidence.

Panel c in Figure 6 shows the average relative impulse response of consumption to a one and two standard deviation increase in the transitory component η_t of income, keeping productivity fixed. We have scaled the results in each case to make them comparable. Consistent with the empirical evidence, we find that the consumption response is relatively smaller in the economy that experiences the larger, transitory increase in income. This is due to the larger increase in the "noise-component" η_t triggering a relative increase in the agent's caution choice $m_t^\star \downarrow$, which, in turn, drives home a relative decrease in expected life-time income.

5.3.3 Pre-cautionary Consumption Choices

We return to the idea that uncertainty about the use of information causes a novel, precautionary motive for consumption. Suppose the agent fixes the weight on productivity $\hat{\beta}_1$ in her expectation of future income y_{t+k}. Because of the linearity of the consumption rule in (5.5), changes in the volatility of income then do not affect current consumption; certainty-equivalence holds. However, despite the linearity of (5.5), once we account for the optimal weight on information $\hat{\beta}_1^\star$ uncertainty once more matters for current consumption.

Corollary 3 and Figure 7 summarize the changes that an increase in the standard deviation of transitory income σ has on consumption, using Proposition 3 and 5.

Corollary 3. *Increases in the standard deviation of transitory income σ cause a decline in current consumption c_t, on average. This is despite the consumption rule in (5.5) being linear.*

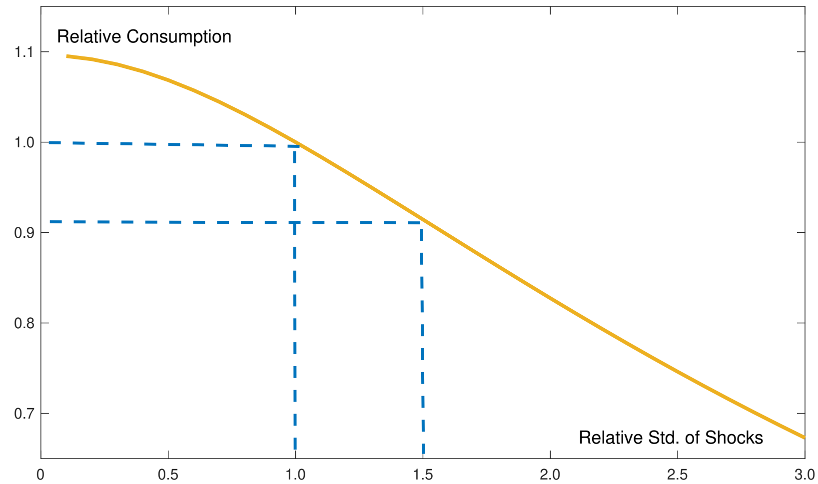
The intuition behind Corollary 3 is simple, and similar to that discussed for Panel c in Figure 6: all else equal, increases in the volatility of transitory income make it more difficult to predict future income based on current productivity. The signal-to-noise ratio of observations is lower. This causes the agent to optimally choose a more cautionary response to information ($m^\star \downarrow$; Proposition 5). As a result, the agent's expectation of life-time income $\{f_t y_{t+k}\}_k$ decreases, on average, which drives down current consumption.

Notice that the agent decreases consumption because of the consequences that noisier income has on her optimal use of information. This contrasts with the case in which the marginal utility of consumption is convex (Kimball, 1990), where agents decrease consumption due to an inherent "dislike" of additional risk. Furthermore, notice that within our framework increases in the volatility of transitory income, such as that which occur during recessions, are always accompanied by decreases in income expectations. By contrast, with knowledge of the optimal weight on information, no such relationship necessarily exists.

Notwithstanding the simplicity of our model, Figure 7 shows that size of this pre-cautionary response can be sizable. Under our benchmark parametrization of income, which matches the time-series properties of economy-wide output (see also Figure 5), a 50 percent increase in the standard deviation of the transitory component of income, on average, causes an 8 percent drop in consumption. This is further assuming that $R = 1$. A 50 percent increase in volatility of the transitory component is roughly in line with the evidence in Storesletten *et al.* (2004) on increases in income risk during recessions. Assuming a CRRA-specification for the utility function instead of the quadratic preferences in (5.1), and knowledge of β_1, the 8 percent fall in consumption, all else equal, translates into a coefficient of risk aversion of around 7 to 8, assuming initial assets equal to zero, $a_1 = 0$.

The results in this subsection have highlighted two implications of uncertainty about the

Figure 7: Pre-cautionary Consumption Choices



Note: The figure illustrates average consumption choices for an agent who has observed income and her own productivity for 13 periods ($t = 13$) and will work for another 20 periods ($T = 33$). The figure shows the relative consumption choice as a function of the relative standard deviation of transitory income shocks σ. Baseline parameters are identical to those in Figure 5 and initial assets $a_1 = 0$.

optimal use of information. First, uncertainty about the use of information not only affects the properties of expectations, but also heightens the persistence of consumption fluctuations and makes them history-dependent. Second, the lack of knowledge of how to best predict income creates a novel, pre-cautionary motive for savings. Because of uncertainty about the optimal weight on information, agents choose to consume less at the start of their life, as well as when faced with larger transitory income shocks. The latter can help account for the size-dependent effects of stimulus checks observed in the data.

6 Final Remarks

In this paper, we have presented a tractable and parsimonious theory of cautious expectations based on the observation that the optimal use of information is often unknown. Our framework formalizes the idea that, to form expectations, agents first need to decide how much weight to attach to the various signals observed about a forecasted variable. The agent-econometrician assumption, in which agents estimate the optimal weight on their own information from the history of past observations, provides a natural departure point to model this idea.

The broad principle emerging from our analysis is that uncertainty about the optimal use of information leads to a cautionary response to prior and new information that exploits a fundamental trade-off between the bias and variance of expectations. The observed bias in survey expectations of professional forecasters combined with their accuracy, as well as the documented co-existence of under- and overreactions to new information, can be viewed as specific manifestations of this mechanism. The structure of caution choices, furthermore, yields new predictions about the time- and state-dependence of expectations.

To illustrate our results, we embedded our framework into a standard consumption-savings problem that lies at the bedrock of modern macroeconomic models. For standard parameters, we have showed that uncertainty about the optimal weight on information in predicting future income leads to an upward-sloping consumption profile and muted, persistent responses to income shocks. In addition, we have demonstrated that increased uncertainty about the best use of information results in a sizable decline in current consumption, even with linear decision rules. We have further discussed how our results can help match evidence on the responsiveness of consumption to the 2001 and 2008 stimulus package, respectively.

In spirit, our analysis has been close those in the rational and behavioral inattention literatures (Maćkowiak *et al.*, 2021; Gabaix, 2017). The central difference is that our model focuses on agents' optimal responses to uncertainty about the *use* of information. The rational and behavioral inattention literatures, by contrast, center on what information agents *choose* to observe. In this sense, our theory provides a complement rather than a substitute to those based on limited attention. Each focuses on a different "stage" of the expectation formation process. Models, such as that analyzed in Section 4, are interesting in this respect because they merge a model of noisy information, due to, for instance, limited attention, with the cautious responses that arise due to uncertainty about the optimal use of information.

One appealing feature of our framework is that it arrives at predictions that are independent of any cost function on beliefs. This is important because, although several tractable models of, for example, attention costs exist, neither accurately matches the mixed evidence on costs faced by households and firms (e.g., Carlsson and Skans, 2012; Caplin and Dean, 2015; Dewan and Neligh, 2020). By contrast, the implied caution choices in our model are derived solely from the economic incentives and technologies that agents face, and are as such not confounded by potential misspecifications in cost functions.

Beyond the analysis in this paper, future work may use our results to shed light on several issues. One natural candidate is the documented history- and state-dependence of expectations, which has been shown to influence economic decisions by households and firms (e.g., Malmendier and Wachter, 2021). These studies often rely on semi-structural or ad-hoc rules about the dependence of individual expectations on the history of observations. This contrasts with the optimal-derived, history-dependent expectations studied above. Cautious expectations may also prove useful for thinking about the excessive risk-premium on risky assets, which have been documented since Mehra and Prescott's (1985) influential contribution. Compte and Postlewaite (2019) discuss how risk premia can arise from a market in which traders learn about a structural parameter related to risk. More can be done to connect a cautious use of information to the preference for safe assets in financial portfolios.[35] Finally, benchmark models of information choice abstract from any uncertainty about the use of signals. By contrast, in reality, it seems signals are often selected based on agents' knowledge about their optimal use. Incorporating uncertainty about the optimal use of information into a standard framework for information choice seems like a natural next step.

[35] Haworth and Gai (2023) show how the framework developed in this paper can also be used to cast light on the trade-offs involved in macroprudential policy.

References

Afrouzi, H., Kwon, S. Y., Landier, A., Ma, Y. and Thesmar, D. (2021). Overreaction and working memory. *National Bureau of Economic Research*.

Anderson, B. D. and Moore, J. B. (1979). Optimal filtering. *Englewood Cliffs*, **21**, 22–95.

Angeletos, G.-M. and Huo, Z. (2021). Myopia and anchoring. *American Economic Review*, **111** (4), 1166–1200.

—, — and Sastry, K. A. (2021). Imperfect macroeconomic expectations: Evidence and theory. *NBER Macroeconomics Annual*, **35** (1), 1–86.

— and Lian, C. (2016). Incomplete information in macroeconomics: Accommodating frictions in coordination. *Handbook of Macroeconomics*, **2**, 1065–1240.

— and Pavan, A. (2007). Efficient Use of Information and Social Value of Information. *Econometrica*, **75** (4), 1103–1142.

Aras, E., Lee, K.-Y., Pananjady, A. and Courtade, T. A. (2019). A family of bayesian cramér-rao bounds, and consequences for log-concave priors. In *2019 IEEE International Symposium on Information Theory (ISIT)*, IEEE, pp. 2699–2703.

Barberis, N., Greenwood, R., Jin, L. and Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, **129** (2), 203–227.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, **2**, 69–186.

Bennett, J. and Owyang, M. (2022). On the relative performance of inflation forecasts. *Federal Reserve Bank of St. Louis Review*.

Bordalo, P., Gennaioli, N., Ma, Y. and Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, **110** (9), 2748–82.

—, — and Shleifer, A. (2017). Memory, attention, and choice. *The Quarterly journal of economics*.

—, — and — (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, **73** (1), 199–227.

Brainard, W. C. (1967). Uncertainty and the effectiveness of policy. *The American Economic Review*, **57** (2), 411–425.

Broer, T. and Kohlhas, A. N. (2022). Forecaster (mis-)behavior. *The Review of Economics and Statistics (forthcoming)*.

Cagan, P. (1956). The monetary dynamics of hyperinflation. *Studies in the Quantity Theory if Money*.

Canova, F. (2011). *Methods for applied macroeconomic research*. Princeton university press.

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, **105** (7), 2183–2203.

Carlsson, M. and Skans, O. N. (2012). Evaluating microfoundations for aggregate price rigidities: evidence from matched firm-level data on product prices and unit labor cost. *American Economic Review*, **102** (4), 1571–95.

Coibion, O. and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, **120** (1), 116–159.

— and — (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, **105** (8), 2644–78.

—, — and Kumar, S. (2018). How do firms form their expectations? new survey evidence. *American Economic Review*, **108** (9), 2671–2713.

Compte, O. and Postlewaite, A. (2019). *Ignorance and uncertainty*. 61, Cambridge University Press.

Cramér, H. (1946). A contribution to the theory of statistical estimation. *Scandinavian Actuarial Journal*, **1946** (1), 85–94.

Croushore, D. D. (1993). Introducing: the survey of professional forecasters. *Business Review-Federal Reserve Bank of Philadelphia*, **6**, 3.

Cutler, D. M., Poterba, J. M. and Summers, L. H. (1990). Speculative dynamics and the role of feedback traders. *American Economic Review*, **80** (2), 63–68.

Da Silveira, R. A., Sung, Y. and Woodford, M. (2022). Optimally imprecise memory and biased forecasts.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, **98** (4), 703–738.

Dewan, A. and Neligh, N. (2020). Estimating information cost functions in models of rational inattention. *Journal of Economic Theory*, p. 105011.

Eldar, Y. C. (2008). Rethinking biased estimation: Improving maximum likelihood and the cramér–rao bound. *Foundations and Trends in Signal Processing*, **1** (4), 305–449.

Elliott, G., Komunjer, I. and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, **6** (1), 122–157.

Evans, G. W. and Honkapohja, S. (2012). *Learning and expectations in macroeconomics*. Princeton University Press.

Farmer, L., Nakamura, E. and Steinsson, J. (2022). Learning about the long run.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In *Handbook of economic forecasting*, vol. 2, Elsevier, pp. 2–56.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J. and Watson, M. W. (2007). Abcs (and ds) of understanding vars. *American economic review*, **97** (3), 1021–1026.

Flynn, J. P. and Sastry, K. (2021). Attention cycles. *MIT mimeo.*

Fréchet, M. (1943). Sur l'extension de certaines évaluations statistiques au cas de petits échantillons. *Revue de l'Institut International de Statistique*, pp. 182–205.

Fuhrer, J. (2018). Intrinsic expectations persistence: Evidence from professional and household survey expectations. *Boston Fed mimeo.*

Fuster, A., Laibson, D. and Mendel, B. (2010). Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives*, **24** (4), 67–84.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, **129** (4), 1661–1710.

— (2017). Behavioral inattention. *Handbook of Behavioral Economics.*

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC press.

Gemmi, L. and Valchev, R. (2022). Public information and survey expectations.

Golan, A., Judge, G. and Miller, D. (1997). *Maximum entropy econometrics: Robust estimation with limited data*. John Wiley and Sons.

Gourinchas, P.-O. and Parker, J. A. (2002). Consumption over the life cycle. *Econometrica*, **70** (1), 47–89.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of Political Economy*, **86** (6), 971–987.

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton University Press.

Harvey, A. C. (1990). *The econometric analysis of time series*. MIT Press.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Haworth, C. and Gai, P. (2023). Macroprudential policymakers with cautious expectations.

Ilut, C. L. and Valchev, R. (2022). Economic agents as imperfect problem solvers. *Quarterly Journal of Economics.*

Johnson, D. S., Parker, J. A. and Souleles, N. S. (2006). Household expenditure and the income tax rebates of 2001. *American Economic Review*, **96** (5), 1589–1610.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological review*, **80** (4), 237.

Kimball, M. S. (1990). Precautionary saving in the small and in the large. *Econometrica*, **58** (1), 53–73.

Kohlhas, A. N. and Walther, A. (2021). *American Economic Review*, **111** (9), 2879–2925.

Kőszegi, B. and Matějka, F. (2020). Choice simplification: A theory of mental budgeting and naive diversification. *The Quarterly Journal of Economics*, **135** (2), 1153–1207.

Kozlowski, J., Veldkamp, L. and Venkateswaran, V. (2020). The tail that wags the economy: Beliefs and persistent stagnation. *Journal of Political Economy*, **128** (8), 2839–2879.

Kwon, S. Y. and Tang, J. (2022). Extreme events and overreaction to news.

Lucas, R. E. J. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, **4** (2), 103–124.

Maćkowiak, B., Matějka, F. and Wiederholt, M. (2021). Rational inattention: A review.

— and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *The American Economic Review*, **99** (3), 769–803.

Malmendier, U. and Wachter, J. A. (2021). Memory of past experiences and economic decisions. *Available at SSRN 4013583*.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the nesw keynesian phillips curve. *The Quarterly Journal of Economics*, **117** (4), 1295–1328.

Marinovic, I., Ottaviani, M. and Sorensen, P. (2013). Forecasters objectives and strategies. *Handbook of economic forecasting*, **2**, 690–720.

Matějka, F. (2016). Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, **83** (3), 1125–1155.

Mitchell, J., Robertson, D. and Wright, S. (2019). R2 bounds for predictive models: what univariate properties tell us about multivariate predictability. *Journal of Business & Economic Statistics*, **37** (4), 681–695.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, **92** (5), 1521–1534.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, pp. 315–335.

Nimark, K. (2008). Dynamic pricing and imperfect common knowledge. *Journal of Monetary Economics*, **55** (2), 365–382.

Onatski, A. and Stock, J. H. (2000). Robust monetary policy under model uncertainty in a small model of the us economy. *Macroeconomic Dynamics*.

Orcutt, G. H. and Winokur Jr, H. S. (1969). First order autoregression: inference, estimation, and prediction. *Econometrica: Journal of the Econometric Society*, pp. 1–14.

Parker, J. A., Souleles, N. S., Johnson, D. S. and McClelland, R. (2013). Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, **103** (6), 2530–53.

Phillips, L. D. and Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of experimental psychology*, **72** (3), 346.

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, **37**, 81–91.

Sargent, T. J. (1987). *Dynamic macroeconomic theory*. Harvard University Press.

— (1993). Bounded rationality in macroeconomics: The arne ryde memorial lectures. *OUP Catalogue*.

Scharfstein, D. S. and Stein, J. C. (1990). Herd behavior and investment. *The American Economic Review*, pp. 465–479.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50** (3), 665–690.

— (2006). Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, **96** (2), 158–163.

Stark, T. (2010). Realistic evaluation of real-time forecasts in the survey of professional forecasters. *Federal Reserve Bank of Philadelphia Special Report*, **1**.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*.

Stock, J. H. and Watson, M. W. (2008). Phillips curve inflation forecasts.

Storesletten, K., Telmer, C. I. and Yaron, A. (2004). Cyclical dynamics in idiosyncratic labor market risk. *Journal of political Economy*, **112** (3), 695–717.

Tanaka, K. (1984). An asymptotic expansion associated with the maximum likelihood estimators in arma models. *Journal of the Royal Statistical Society: Series B (Methodological)*, **46** (1), 58–67.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58** (1), 267–288.

Tikhonov, A. N. and Arsenin, V. Y. (1977). Solutions of ill-posed problems. *American Mathematical Society*, **1** (30), 487.

Van Trees, H. L. and Bell, K. L. (1968). *Detection estimation and modulation theory*. Wiley.

Varian, H. R. (1975). A bayesian approach to real estate assessment. *Studies in Bayesian econometric and statistics in Honor of Leonard J. Savage*, pp. 195–208.

Woodford, M. (2002). *Imperfect Common Knowledge and the Effects of Monetary Policy*. Nber working papers, Department of Economics, Columbia University.

Zarnowitz, V. (1985). Rational expectations and macroeconomic forecasts. *Journal of Business & Economic Statistics*, **3** (4), 293–311.

Appendix to

"Cautious Expectations"

Alexandre N. Kohlhas and Donald Robertson
University of Oxford and University of Cambridge

February, 2024

Contents of Appendix:

This *Appendix* details extensions and proofs of the models analyzed in the paper.

A Proofs and Derivations

A.1 Optimal Expectations

Proof of Proposition 1: The proof follows van Tries (1968) but extends the result to the vector case. For ease of notation, we suppress that all moments are conditional on x^t. Thus, we for example write $\mathbb{E}[V]$ instead of $\mathbb{E}\left[V \mid x^t\right]$ for any random vector V.

Let z be given by $z = \left(x^t, y^t\right)$. By the definition of the bias function, we have that

$$\beta + b(\beta) = \mathbb{E}\left[\hat{\beta}\right] = \int \hat{\beta}\left(z\right) f\left(z;\beta\right) dz. \tag{A1}$$

Let $\nabla f(z;\beta)$ denote the partial derivate of the joint probability density function with respect to β. Differentiating both sides of (A1) provides us with

$$\begin{aligned} I_n + D\left(\beta\right) &= \int \hat{\beta}\left(z\right) \nabla f(z;\beta)' dz \\ &= \int \hat{\beta}\left(z\right) \nabla f(z;\beta)' \frac{f(z;\beta)}{f(z;\beta)} dz. \end{aligned}$$

Notice that $\nabla f(z;\beta)/f(z;\beta)$ is equal to the partial derivate of the log-likelihood function $\mathcal{L}\left(z;\beta\right) \equiv \log f\left(z;\beta\right)$ with respect to β. Thus,

$$I_n + D\left(\beta\right) = \mathbb{E}\left[\hat{\beta}(z) \nabla \mathcal{L}(z;\beta)'\right]. \tag{A2}$$

Now, take the fact that

$$1 = \int f\left(z;\beta\right) dz$$

and differentiate both sides to arrive at

$$\mathbf{0} = \int \nabla f(z;\beta)' dz = \int \nabla f(z;\beta)' \frac{f(z;\beta)}{f(z;\beta)} dz = \mathbb{E}\left[\nabla \mathcal{L}(z;\beta)'\right].$$

Pre-multiplying this expression with $\mathbb{E}\left[\hat{\beta}(z)\right]$ yields

$$\mathbf{0} = \mathbb{E}\left[\hat{\beta}(z)\right] \mathbb{E}\left[\nabla \mathcal{L}(z;\beta)'\right]. \tag{A3}$$

Deducting (A2) from (A3) now provides us with

$$I_n + D\left(\beta\right) = \mathbb{E}\left[\left(\hat{\beta}(z) - \mathbb{E}\left[\hat{\beta}(z)\right]\right) \nabla \mathcal{L}(z;\beta)'\right]. \tag{A4}$$

The vector case of the Schwarz inequality shows that[1]

$$\begin{aligned}\mathbb{V}\left[\hat{\beta}\left(z\right)\right] &\succeq \mathbb{E}\left[\left(\hat{\beta}(z)-\mathbb{E}\left[\hat{\beta}(z)\right]\right)\nabla\mathcal{L}(z;\beta)'\right] \\ &\times \mathbb{E}\left[\nabla\mathcal{L}(z;\beta)\nabla\mathcal{L}(z;\beta)'\right]^{-1} \\ &\times \mathbb{E}\left[\left(\hat{\beta}(z)-\mathbb{E}\left[\hat{\beta}(z)\right]\right)\nabla\mathcal{L}(z;\beta)'\right]'.\end{aligned} \tag{A5}$$

Inserting (A4) into (A5) then completes the proof. □

Proof of Proposition 2: Let $\hat{\beta} = M\hat{\beta}^{ls} \equiv \left(I_n + K\right)\hat{\beta}^{ls}$ be an estimator of β, with bias function $b(\beta) = \mathbb{E}\left[\hat{\beta} \mid x^t\right] - \beta = K\beta$. Then $\mathbb{V}\left[\hat{\beta} \mid x^t\right]$ is bounded below by the FCR lower-bound:

$$\mathbb{V}\left[\hat{\beta} \mid x^t\right] \succeq \left[I_n + K\right](X'X)^{-1}\sigma^2\left[I_n + K\right]',$$

where $D\left(\beta\right) = db\left(\beta\right)/d\beta = K$, and we have used that

$$J = \sigma^{-2}\left(X'X\right), \quad X \equiv \left[\begin{array}{cccc} x_1 & x_2 & \ldots & x_n \end{array}\right].$$

Now notice that, because $\hat{\beta}^{ls}$ attains the unbiased FCR lower-bound, we have that

$$\mathbb{V}\left[\hat{\beta}^{ls} \mid x^t\right] = (X'X)^{-1}\sigma^2.$$

Thus,

$$\mathbb{V}\left[\hat{\beta} \mid x^t\right] = \left[I_n + K\right](X'X)^{-1}\sigma^2\left[I_n + K\right]'.$$

□

Proof of Proposition 3: We consider estimators $\hat{\beta} = M\hat{\beta}^{ls}$, where $M \in \mathbb{R}^{n\times n}$. Inserting such a scaled estimator into (3.1) and (3.2) shows that

$$\mathcal{U} = -\frac{1}{2}\left(\mathbb{V}\left[\eta_{t+k}\right] + \mathbb{E}_x x_t'\left[MJ^{-1}M' + (I_n - M)\beta\beta'(I_n - M)'\right]x_t\right).$$

The agent's optimization problem for all x_t thus becomes

$$\max_{M\in\mathbb{R}^{n\times n}} -\frac{1}{2}\mathbb{E}_x x_t'\left[MJ^{-1}M' + (I_n - M)\beta\beta'(I_n - M)'\right]x_t. \tag{A6}$$

The sufficient first-order condition is

$$\partial\mathcal{U}/\partial M = 0: \quad \mathbb{E}_x\left\{\left[\left(J^{-1} + \beta\beta'\right)M^{\star\prime} - \beta\beta'\right]x_t x_t'\right\} = 0, \tag{A7}$$

Thus, if (A7) has to hold for all x_t,

$$M^\star = \beta\beta'\left(J^{-1} + \beta\beta'\right) = I_n - J^{-1}\left(J^{-1} + \beta\beta'\right)^{-1}, \tag{A8}$$

[1] We use the notation that, for two matrices A and B, $A \succeq B$ means that $A - B$ is positive (semi-)definite.

where the final equality follows from an application of the Woodbury Identity.
The Woodbury Identity further allows us to write the previous expression as

$$M^{\star} = \frac{1}{1+\beta' J\beta}\beta\beta' J. \tag{A9}$$

It now follows from (A9) and $\hat{\beta}^{\star} = M^{\star}\hat{\beta}^{ls}$, where $\mathbb{E}\left[\hat{\beta}^{ls} \mid x^t\right] = \beta$, that

$$\mathbb{E}\left[\hat{\beta}^{\star} \mid x^t\right] = \frac{1}{1+\beta' J\beta}\beta\beta' J\mathbb{E}\left[\hat{\beta}^{ls} \mid x^t\right] = m^{\star}\beta,$$

where $m^{\star} \equiv (1+\beta' J\beta)^{-1}\beta' J\beta \in [0,1)$ since J is positive semi-definite.
A few tedious but simple matrix derivations then show that

$$\mathbb{E}\left[\hat{\beta}_j^{\star} \mid x^t\right] = m^{\star}\beta_j, \quad m^{\star} = \frac{\sum_i \sum_k \beta_i\beta_k\left(\frac{x_i' x_k}{\sigma^2}\right)}{1+\sum_i \sum_k \beta_i\beta_k\left(\frac{x_i' x_k}{\sigma^2}\right)} \in [0,1). \tag{A10}$$

Finally, notice that $m^{\star}$ in (A10) can alternatively be written as

$$m^{\star} = \frac{\frac{1}{\sigma^2}(X\beta)'(X\beta)}{1+\frac{1}{\sigma^2}(X\beta)'(X\beta)} = \frac{\frac{1}{\sigma^2}\hat{y}_{[t]}'\hat{y}_{[t]}}{1+\frac{1}{\sigma^2}\hat{y}_{[t]}'\hat{y}_{[t]}},$$

where $\hat{y}_{[t]}$ denotes the true vector of fitted values based on period time t information. □

A.2 A Partial Equivalence Result

Proof of Proposition 4: The proof proceeds in four steps. First, we derive the rational expectation when the agent observes x_t and knows the true value of β. Then, we turn to the noisy-information case, in which the agent instead observes the noisy signal z_{it}, before detailing the corresponding optimal expectation $f_t^{\star}y_{t+k}$. The fourth and final steps shows that a noise choice q exist such that we can equate the average value of the former with the expected value of the latter for each value of x_t.

Step 1: The informed expectation, assuming knowledge of β, is

$$\begin{aligned} y_{t+k|t} &= \mathbb{E}\left[y_{t+k} \mid x_t; \beta\right] \\ &= \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_n x_{nt}, \end{aligned}$$

Thus, for a certain value of σ, we can write

$$y_{t+k} = x_t'\beta + \eta_{t+k}, \quad \eta_{t+k} \sim \mathcal{N}\left(0, \sigma^2\right).$$

Step 2: The noisy rational expectation is

$$\mathbb{E}\left[y_{t+k} \mid z_{it}; \hat{m}\right] = \hat{m} z_{it},$$

where

$$\hat{m} \equiv \frac{q^{-2}}{q^{-2} + \mathbb{V}ar\left[x_t'\beta\right]}. \tag{A11}$$

The *average noisy rational expectation* is therefore:

$$\int_0^1 \mathbb{E}\left[y_{t+k} \mid z_{it}\right] di = \hat{m}\mathbb{E}\left[y_{t+k} \mid x_t; \beta\right] = \hat{m}x_t'\beta.$$

Step 3: The optimal expectation is, in contrast, equal to

$$f_t^{\star} y_{t+k} = \hat{\beta}_1^{\star} x_{1t} + \hat{\beta}_2^{\star} x_{2t} + \ldots + \hat{\beta}_n^{\star} x_{nt},$$

where

$$\mathbb{E}\left[\hat{\beta}_j^{\star} \mid x_t\right] = m^{\star}\beta_j.$$

Thus,

$$\mathbb{E}\left[f_t^{\star} y_{t+k} \mid x_t\right] = m^{\star} x_t'\beta.$$

Step 4: We will show there $\exists q \in \mathbb{R}_+ : \int_0^1 \mathbb{E}\left[y_{t+k} \mid z_{it}\right] di = \mathbb{E}\left[f_t^{\star} y_{t+k} \mid x_t\right]$ for all x_t. This, however, follows from (A11) since $\hat{m}$ can admit all values in the range between zero and one. □

A.3 Comparative Statics

Proof of Proposition 5: The proof has two parts.

Part (i): The partial derivative of $m^{\star}$ with respect to $\sum_{l=1}^{l=t}\left(\frac{x_{jl}}{\sigma}\right)^2$ is

$$\frac{\partial m^{\star}}{\partial \sum_{l=1}^{l=t}\left(\frac{x_{jl}}{\sigma}\right)^2} = \beta_j^2 \times \frac{1}{\left[1 + \sum_i \sum_k \beta_i \beta_k \left(\frac{x_i' x_k}{\sigma^2}\right)\right]^2} \geq 0.$$

Part (ii): The partial derivative of $m^{\star}$ with respect to $\sum_{l=1}^{l=t}\left(\frac{x_{it}x_{jt}}{\sigma}\right)^2$ is

$$\frac{\partial m^{\star}}{\partial \sum_{l=1}^{l=t}\left(\frac{x_{it}x_{jt}}{\sigma}\right)^2} = 2 \times \frac{\beta_i \beta_j}{\left[1 + \sum_i \sum_k \beta_i \beta_k \left(\frac{x_i' x_k}{\sigma^2}\right)\right]^2}.$$

This completes the proof. □

Proof of Proposition 6: The comparative statics can be shown as follows:

Part (i): If $\sum_t x_{it}x_{jt} = 0$ for all $i \neq j$, then $m^{\star}$ in (A10) becomes

$$m^{\star} = \frac{\sum_{i=1}^{i=n} \beta_i^2 \sum_{l=1}^{l=t}\left(\frac{x_{il}}{\sigma}\right)^2}{1 + \sum_{i=1}^{i=n} \beta_i^2 \sum_{l=1}^{l=t}\left(\frac{x_{il}}{\sigma}\right)^2}.$$

Because both $\beta_i^2 \geq 0$ and $\sum_{l=1}^{l=t}\left(\frac{x_{il}}{\sigma}\right)^2 \geq 0$, this shows that $m^{\star}$ increases in n.

Part (ii): Consider the equation for $m^{\star}$ in (A10). The contribution of the nth signal to the numerator and denominator of this expression is

$$a_n \equiv \beta_n^2 \sum_t x_{nt}^2/\sigma^2 + \beta_n \sum_{i \neq n} 2\beta_i \sum_t x_{nt} x_{jt}.$$

Consequently, if $a_n < 0$, $m^\star$ decreases with the addition of the nth signal.

Part (iii): Finally, notice that $m^\star$ in (A10) based upon the first t observations equals

$$m^\star_{[t]} = \frac{\frac{1}{\sigma^2}\left(X\beta\right)'\left(X\beta\right)}{1 + \frac{1}{\sigma^2}\left(X\beta\right)'\left(X\beta\right)} = \frac{\frac{1}{\sigma^2}\hat{y}'_{[t]}\hat{y}_{[t]}}{1 + \frac{1}{\sigma^2}\hat{y}'_{[t]}\hat{y}_{[t]}},$$

where $\hat{y}_{[t]}$ denotes the vector of fitted values based on period time t information.
But this shows that $m^\star_{[t+1]}$ is equal to

$$m^\star_{[t+1]} = \frac{\frac{1}{\sigma^2}\hat{y}'_{[t]}\hat{y}_{[t]} + \frac{1}{\sigma^2}\left(\hat{y}_{t+1}\right)^2}{1 + \frac{1}{\sigma^2}\hat{y}'_{[t]}\hat{y}_{[t]} + \frac{1}{\sigma^2}\left(\hat{y}_{t+1}\right)^2} \geq m^\star_{[t]}.$$

□

B Extensions and Further Results

B.1 Iterative and Robust Estimation

Proposition B.1. ***(An Iterative Estimator)*** Consider the estimator $\hat{\beta}^{\star}$ in (3.7), and let

$$\hat{\beta}^{\star}_{[k]} = \left[I_N - J^{-1} \left(J^{-1} + \hat{\beta}^{\star}_{[k-1]} \hat{\beta}^{\star}_{[k-1]} \right)^{-1} \right] \hat{\beta}^{\star}_{[k-1]}, \tag{A12}$$

where $\hat{\beta}^{\star}_{[0]} = \hat{\beta}^{ls}$. Then, the fixed-point of (A12) dominates $\hat{\beta}^{ls}$ in terms of agents' payoffs for all value of $\beta \in \mathbb{R}^n$ if the effective dimension $d \geq 4$, where $d \equiv tr\,(Q) / \lambda_{max}\,(Q)$ and $Q \equiv (X'X)^{-1}$.[2]

Proof of Proposition B.1: The proof follows the steps for Theorem 2 in Eldar (2008).

Proposition B.2. ***(A Robust Min-Max Estimator)*** Let $\beta_j \in \left[\underline{\beta}_j, \bar{\beta}_j \right]$ and consider the optimal estimator $\hat{\beta}^{\star}$ and the associated caution choice $m^{\star}$. The estimator that replaces β_j in the expressions for $\hat{\beta}^{\star}$ and $m^{\star}$ with either $\underline{\beta}_j$ or $\bar{\beta}_j$, depending on which yields the lowest utility, strictly dominates the least-squares (no-caution) alternative for all values of β.

Proof of Proposition B.2: For any estimator that satisfies Proposition 2, we have that[3]

$$mse\left(\hat{\beta}\right) = MJ^{-1}M' + (I_n - M)\,\beta\beta'\left(I_n - M'\right), \tag{A13}$$

where J denotes the information matrix. We want to minimize the quadratic form $x_t' mse\left(\hat{\beta}\right) x_t$, because it is what determines utility losses in (3.1). The problem is that the optimal value $M^{\star}$ depends on unknown parameters. A robust min-max approach to this problem for all x_t is

$$M^{robust} = \arg\min_{M} \max_{\beta_j \in \left(\underline{\beta}_j, \bar{\beta}_j\right)} x_t' mse\left(\hat{\beta}\right) x_t. \tag{A14}$$

Because of the quadratic nature of (A13), we know that a solution to (A14) dominates the least-squares outcome $M = I_n$. Moreover, a solution to this problem can always be found.

Notice that (A14) can also be written as

$$M^{robust} = \arg\min_{M} \left\{ x_t M J^{-1} M' x_t' + \max_{\beta_j \in \left(\underline{\beta}_j, \bar{\beta}_j\right)} x_t' \left(I_n - M\right)\beta\beta'\left(I_n - M'\right) x_t \right\}.$$

where the term inside the max-operator clearly either always has a positive (or negative) gradient. So the solution to this problem must line on the edges of the constraint. The worst-case β_j has to be either $\underline{\beta}_j$ or $\bar{\beta}_j$. Inserting the worst-case β into (3.7) and (3.8) completes the proof.

Finally, notice that a natural constraint $\beta_j \in (0, 1)$ for all j arises, for example, in the tracking problems studied in Section 4.2. $\square$

[2] $\lambda_{\max}(Q)$ here denotes the maximum eigenvalue of Q. Furthermore, we note that X once more equals $X \equiv \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}$, consistent with the use of X in the paper and Appendix A.

[3] Without loss of generality, we here treat the signal vector x^t as fixed.

B.2 A Non-linear and a Non-Gaussian Example

This appendix shows how our main results extend beyond the linear-Gaussian case.

A Non-linear Example: Consider the setup from Section 2, but suppose that y_{t+k} is independent over time and uniformly distributed on $[0, \beta]$, where β is an unknown positive parameter.

The probability density function is

$$f(y) = \begin{cases} \frac{1}{\beta} & 0 \leq y \leq \beta \\ 0 & \text{elsewhere.} \end{cases}$$

Agents observe t realizations of y_t, so that their information set is $\Omega_t = \{y_\tau\}_{\tau=1}^{\tau=t}$.

It follows that

$$\mathbb{E}\left[y_t\right] = \frac{\beta}{2}, \quad \mathbb{V}[y] = \frac{\beta^2}{12}.$$

The *minimum variance unbiased estimator* for the uniform case is

$$\hat{\beta}^{mvu} = (1 + 1/T)\, y_{max}, \quad y_{max} = \max_{\tau} y_\tau. \tag{A15}$$

We note that the estimator in (A15) is non-linear in the observations of y_t.

To derive an estimator that produces more accurate forecasts, consider the alternative estimator:

$$\hat{\beta} = m\hat{\beta}^{mvu}, \quad m \in \mathbb{R}. \tag{A16}$$

As in the main test, the estimator in (A16) is a linear function of $\hat{\beta}^{mvu}$, and hence has a linear bias. It follows from these two features (as in Proposition 2) that (A16) attains the lower-bound on the variance of any biased estimator. Following the same steps as in Section 3 shows that

$$\hat{\beta}^{\star} = \frac{t+2}{t+1} y_{max}, \quad y_{max} = \max_{\tau} x_\tau \tag{A17}$$

has a smaller mean-squared-error than any other estimator of β. We conclude that $f_t^{\star} y_{t+1} = \frac{1}{2}\hat{\beta}^{\star}$ produces the most accurate forecast.

Finally, notice that $m^{\star}$ is independent of the realizations of the signal in Ω_t. The estimator in (A17) thus provides an example of an estimator that is directly implementable.

We close this example with two implications of the optimal adjustment in (A17) that mirror those in Section 4 of the main text. First, the optimal estimator results in biased forecasts:

$$\mathbb{E}\left[y_{t+1} - f_t^{\star} y_{t+1}\right] = \frac{1}{2}\left(\beta - \mathbb{E}\hat{\beta}^{\star}\right) = \frac{1}{2}\frac{1}{(1+t)^2}\theta > 0.$$

Second, the optimal estimator results in errors that can be predicted based on observed signals:

$$\mathbb{P}\left[y_{t+1} - f_t^{\star} y_{t+1} \mid y_t\right] = y_{t+1} - \mathbb{P}\left[\frac{1}{2}\frac{t+2}{t+1} y_{max} \mid y_t\right] = \left(1 - \frac{t(t+2)}{(1+t)^2}\right) y_t = \frac{1}{(1+t)^2} y_t,$$

where $\mathbb{P}[z_t \mid y_t]$ denotes the linear projection of z_t onto y_t and $(1+t)^{-2} > 0$.

A Non-Gaussian Example: Consider once more the setup from Section 2, but suppose that y_{t+k} is iid across time and Bernoulli distributed with probability of success $\beta \in \left(0, \frac{1}{2}\right]$.[4]

We can write this example in a linear framework:

$$y_{t+k} = \beta + \eta_{t+k},$$

where

$$\eta_{t+k} = \begin{cases} 1-\beta & \text{with prob. } \beta \\ -\beta & \text{with prob. } 1-\beta \end{cases}$$

so that $\mathbb{E}[\eta_{t+k}] = 0$. An agent's information set is $\Omega_t = \{y_\tau\}_{\tau=1}^{\tau=t}$

The *maximum likelihood estimator* of β, which here attains the *unbiased* FCR bound is:

$$\hat{\beta}^{ml} = \bar{y}_t, \quad \bar{y}_t \equiv \frac{1}{t}\sum_{\tau=1}^{\tau=t} y_t.$$

Trivially, $\mathbb{E}\left[\hat{\beta}^{ml}\right] = \beta$ and $\mathbb{V}\left[\hat{\beta}^{ml}\right] = \beta(1-\beta)\frac{1}{t}$.

Now, consider instead the alternative estimator:

$$\hat{\beta} = m\hat{\beta}^{ml}, \quad m \in \mathbb{R}. \tag{A18}$$

Notice that this alternative estimator is a linear function of its maximum likelihood counterpart, and hence has a linear bias. Once more, because of these two features (as in Proposition 2), the estimator in (A18) attains the *biased* FCR lower-bound. Furthermore, we have that

$$b\left(\hat{\beta}\right) = (m-1)\,\beta, \quad \mathbb{V}\left[\hat{\beta}\right] = m^2\beta(1-\beta)\frac{1}{t}$$

Thus, the mean-squared error of $\hat{\beta}$ straightforwardly becomes

$$mse\left(\hat{\beta}\right) = m^2\beta(1-\beta)\frac{1}{t} + (m-1)^2\beta^2.$$

We conclude that $m^\star = \frac{\beta}{\beta+(1-\beta)t^{-1}} \in (0,1)$.

As in the linear-normal case studied in the body of this paper, down-weighing the maximum likelihood estimator is optimal. It once more allows agents to increase the accuracy of their forecasts by trading-off, on the one hand, having a lower variance of their forecasts versus, on the other hand, having a larger bias. Finally, we note that because $\partial mse\left(\hat{\beta}_{|m=m^\star}\right)/\partial\beta > 0$, the feasible estimator that replaces β with the "worst-case" outcome $\beta = \frac{1}{2}$ in our expression for $m^\star$ improves upon the maximum likelihood estimator over the entire range for the true parameter β. In this case, $m^\star_{\text{feasible}} = \frac{t}{1+t}$.

[4]Notice that we only need to consider the $\beta \leq \frac{1}{2}$ case since with symmetric preferences we are free to re-define success and failure.

B.3 Component-based Expectations and Bias

Expectations and Shrinkage: Suppose y_{t+k} adopts the component-based structure,

$$y_{t+k} = z_{1t+k} + z_{2t+k} \tag{A19}$$

$$z_{1t+k} = \beta_1 x_{1t} + \eta_{1t+k}, \quad z_{2t+k} = \beta_2 x_{1t} + \beta_3 x_{2t} + \eta_{2t+k}, \tag{A20}$$

where $\eta_{jt+k} \sim \mathcal{N}\left(0, \sigma_j^2\right)$ with $\mathbb{E}\left[\eta_{1t+k}\eta_{2t+k}\right] = 0$.

An agent's optimal (component-based) expectation of y_{t+k} is

$$f_t^{\star,\mathrm{comp}} y_{t+k} \quad = \quad f_t^{\star,\mathrm{comp}} z_{1t+k} + f_t^{\star,\mathrm{comp}} z_{2t+k},$$

where, using our results from Section 3 (Proposition 3), we find that

$$\mathbb{E}\left[\hat{\beta}_1^\star \mid X\right] = m_1^\star \beta_1, \quad \mathbb{E}\left[\hat{\beta}_j^\star \mid X\right] = m_2^\star \beta_j, \quad j = 2, 3. \tag{A21}$$

This, in turn, shows that

$$\begin{aligned} \mathbb{E}\left[f_t^{\mathrm{comp}} y_{t+k} \mid X\right] &= m_1^\star \beta_1 x_{1t} + m_2^\star \left(\beta_2 x_{1t} + \beta_3 x_{2t}\right) \\ &= \left(m_1^\star \beta_1 + m_2^\star \beta_2\right) x_{1t} + m_2^\star \beta_3 x_{2t}. \end{aligned}$$

Now, let δ_j denote the weight on x_{jt} in the reduced-form for y_{t+k}:

$$\begin{aligned} y_{t+k} &= \left(\beta_1 + \beta_2\right) x_{1t} + \beta_3 x_{2t} + \eta_{1t+k} + \eta_{2t+k} \\ &\equiv \delta_1 x_{1t} + \delta_2 x_{2t} + \epsilon_{t+k}, \quad \epsilon_{t+k} \equiv \eta_{1t+k} + \eta_{2t+k}. \end{aligned} \tag{A22}$$

Combined, (A21) and (A22) show that

$$\mathbb{E}\left[f_t^{\star,\mathrm{comp}} y_{t+k} \mid X\right] = \tilde{m}_1^\star \delta_1 x_{1t} + \tilde{m}_2^\star \delta_2 x_{2t}, \tag{A23}$$

where

$$\tilde{m}_1^\star \equiv \frac{m_1 \beta_1 + m_2 \beta_2}{\beta_1 + \beta_2}, \quad \tilde{m}_2^\star \equiv m_2,$$

and we conclude that the optimal caution choice on x_{1t} and x_{2t} is generically different.

Bias of Expectations in the General Case: Consider an agent's error at t,

$$e_t = y_{t+k} - f_t^{\star,\mathrm{comp}} y_{t+k}.$$

The expected value of the error is

$$\mathbb{E}\left[e_t\right] = \mathbb{E}_x\left[\mathbb{E}\left[e_t \mid X\right]\right] = \mathbb{E}_x \sum_j \left(1 - \tilde{m}_j^\star\right) \delta_j x_{jt}. \tag{A24}$$

We note that in the baseline case, where $\tilde{m}_j^\star = m^\star$ for all j, this expression simplifies to

$$\mathbb{E}\left[e_t\right] = \mathbb{E}_x \sum_j \left(1 - m^\star\right) \delta_j x_{jt}. \tag{A25}$$

Equation (A24) shows that an agent's expectations can be consistent with both over- and under-optimism, depending on whether the sufficient statistic $\mathbb{E}_x \sum_j \left(1 - \tilde{m}_j^\star\right) \delta_j x_{jt}$ is negative or positive. In the special case in which x_{jt} is independently and identically distributed across time for all j, this statistic collapses to $\sum_j \left(1 - \mathbb{E}_x \tilde{m}_j^\star\right) \delta_j \mathbb{E}_x x_{jt} \lessgtr 0$. This condition is clearly negative when agents exhibit relatively more caution towards signals that decrease their expectations, on average. When agents exhibit caution towards signals that, on average, decrease their expectations, agents forecasts mechanically become higher. In effect, by being cautious towards such signals, agents place a smaller weight on the dampening effect that comes from their observation. As a result, agents tend to be more optimistic than a corresponding econometrician about the future. Equation (A24) formalizes this intuition for the general case in which agents shrink signals differentially. These results further straightforwardly extend to the $n > 2$-case.

Finally, notice that the results in this subsection depend crucially on how y_{t+k} is "signed". Consider the baseline case, where $\tilde{m}_j^\star = m^\star$ for all j, and suppose $m^\star$ is held fixed and $\mathbb{E}\left[y_{t+k}\right] > 0$. It then follows that

$$\mathbb{E}\left[y_{t+k} - f_t^{\star,comp} y_{t+k}\right] = \mathbb{E}\left[y_{t+k}\right] - m^\star \mathbb{E}\left[\mathbb{E}\left(y_{t+k} \mid X\right)\right] = (1 - m^\star)\mathbb{E}\left[y_{t+k}\right] > 0. \tag{A26}$$

But now suppose that instead of forecasting y_{t+k}, the agent forecasts $q_{t+k} = -y_{t+k}$.

The resulting bias in forecasts is

$$\mathbb{E}\left[q_{t+k} - f_t^{\star,comp} q_{t+k}\right] = \mathbb{E}\left[q_{t+k}\right] - m^\star \mathbb{E}\left[\mathbb{E}\left(q_{t+k} \mid X\right)\right] = (1 - m^\star)\mathbb{E}\left[q_{t+k}\right] < 0, \tag{A27}$$

where $m^\star \in (0, 1)$ is identical to that in (A26) and $\mathbb{E}\left[q_{t+k}\right] < 0$.

B.4 The Law of Iterated Expectations

Proposition B.3. *The agent's optimal expectations $f_t^\star y_{t+k}$ satisfy a "Law of Iterated Expectations": $f_t^\star f_{t+1}^\star y_{t+k} = f_t^\star y_{t+k}$, where $k \geq 1$.*

Proof of Proposition B.3: The proof follows most directly from the equivalence of an agent's optimal expectations when $n = 1$ to those of conditional expectations based upon a Gaussian prior, $f_t^\star y_{t+k} = \mathbb{E}\left[y_{t+k} \mid x^t, y^t, p_\beta\right]$, where $\beta \sim \mathcal{N}\left(0, c^2\right) = p_\beta$. Below, we outline a more constructive proof. We structure the proof in two steps. The first step provides the proof for $n = 1$. The second step shows that we can generalize this case to $n \in \mathbb{Z}_+$.

Step 1: The $n = 1$ case.

Consider the single variable case

$$y_{t+k} = \beta_{1[k]} x_{1t} + \eta_{t+k}.$$

where we make it explicit that $\beta_{1[k]}$ pertains to the $k-$horizon forecast. Proposition 3 shows:

$$\hat{\beta}^{\star}_{1[k]} = \frac{\sum_{\tau} x^2_{1\tau}}{\left(\beta_{1[k]}/\sigma_{[k]}\right)^{-2} + \sum_{\tau} x^2_{1\tau}} \beta^{ls}_{1[k]} = \left(\sum_{\tau} x^2_{1\tau} + \left(\beta_{1[k]}/\sigma_{[k]}\right)^{-2}\right)^{-1} \left(\sum_{\tau} x_{1\tau} y_{\tau}\right), \tag{A28}$$

which is part of the Tikhonov class of estimators. As a result, it solves the penalized $L2$-problem:

$$\hat{\beta}^{\star}_{1[k]} = \arg\min_{\beta_{1[k]}} (Y_{[k]} - X_{[k]}\beta_{1[k]})'(Y_{[k]} - X_{[k]}\beta_{1[k]}) + \left(\beta_{1[k]}/\sigma_{[k]}\right)^{-2} \beta^2_{1[k]},$$

where

$$X_{[k]} = \left[\begin{array}{cccc} x_{11} & x_{12} & \dots & x_{1t-k} \end{array}\right]', \quad Y_{[k]} = \left[\begin{array}{cccc} y_{k+1} & y_{k+2} & \dots & y_t \end{array}\right]'.$$

We will show that we can also write this problem as the solution to a standard least-squares problem. Hence, we will show that optimal expectations when $n = 1$ correspond to a type of linear projection. To start, define

$$X^{\star}_{[k]} \equiv \left[\begin{array}{c} \left(\beta_{1[k]}/\sigma_{[k]}\right)^{-2} \\ X_{[k]} \end{array}\right], \quad Y^{\star}_{[k]} \equiv \left[\begin{array}{c} 0 \\ Y_{[k]} \end{array}\right].$$

Then, $\hat{\beta}^{\star}_{1[k]}$ also solves the least-squares problem

$$\hat{\beta}^{\star}_{1[k]} = \arg\min_{\beta_{1[k]}} (Y^{\star}_{[k]} - X^{\star}_{[k]}\beta_{1[k]})'(Y^{\star}_{[k]} - X^{\star}_{[k]}\beta_{1[k]}).$$

But this means we can write the optimal expectation of the fundamental y_{t+k} at time t as the first out-of-sample prediction of the linear projection $\mathcal{P}$ of $Y^{\star}_{[k]}$ onto $X^{\star}_{[k]}$, $f^{\star}_t y_{t+k} = \mathcal{P}_{X^{\star}_{[k]}} y^{\star}_{t+k}$. It now follows from the *Law of Iterated Projections (*Brockwell and Davis, 1991*)* that

$$f^{\star}_t f^{\star}_{t+1} y_{t+k} = \mathcal{P}_{X^{\star}_{1:t[k]}} \mathcal{P}_{X^{\star}_{1:t+1[k-1]}} y^{\star}_{t+k} = \mathcal{P}_{X^{\star}_{1:t[k]}} y^{\star}_{t+k} = f^{\star}_t y_{t+k},$$

where we use subscripts to also keep a track of the observations that forecasts are based on.

Step 2: The $n \geq 1$ case.

The $n = 1$ case suffices to show the result for any $n \in \mathbb{Z}_+$. The reason is as follows.

To start, let z_t be defined by the relationship

$$1 \times z_t \equiv x'_t \beta_{[k]}$$

so that we can re-state equation (2.2) as

$$y_{t+k} = \delta_{[k]} \times z_t + \eta_{t+k}, \quad \delta_{[k]} = 1. \tag{A29}$$

We will now show that expectations based on (A29) are identical to those based on (2.2).

Consider the vector of fitted values $X_{[k]}\hat{\beta}^{\star}_{[k]}$ based upon equation (2.2) and information until period-t, where $X_{[k]} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}$. A few simple matrix derivations shows that

$$X_{[k]}\hat{\beta}^{\star}_{[k]} = X_{[k]}\beta_{[k]}\left(m^{\star}_{[k]} + \frac{\beta'_{[k]}X'_{[k]}\eta}{\sigma^2_{[k]} + \beta'_{[k]}X'_{[k]}X_{[k]}\beta_{[k]}}\right), \tag{A30}$$

where we have used our results in Proposition 3. However, the expression in (A30) is identical to the vector of fitted values $Z_{[k]}\hat{\delta}^{\star}_{[k]}$, where $Z_{[k]} = \begin{bmatrix} z_1 & z_2 & \ldots & z_{t-k} \end{bmatrix}'$:

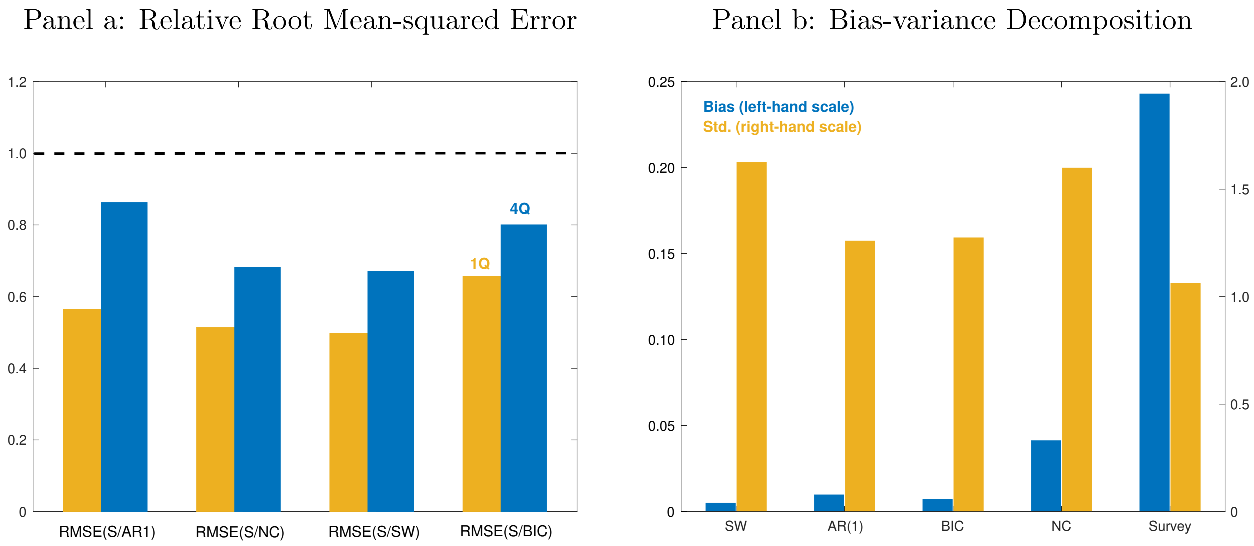
$$Z_{[k]}\hat{\delta}^{\star}_{[k]} = X_{[k]}\beta_{[k]} \times \hat{\delta}^{\star}_{[k]} = X_{[k]}\beta_{[k]}\left(m^{\star}_{[k]} + \frac{\beta'_{[k]}X'_{[k]}\eta}{\sigma^2_{[k]} + \beta'_{[k]}X'_{[k]}X_{[k]}\beta_{[k]}}\right) = X_{[k]}\hat{\beta}^{\star}_{[k]}.$$

Because the entire vector of fitted values is the same, so too is the first out-of-sample prediction, $f^{\star}_t y_{t+k}$. We conclude that $f^{\star}_t y_{t+k}$ using (2.2) with $n \in \mathbb{Z}_+$ is identical to $f^{\star}_t y_{t+k}$ based upon (A29) with $n = 1$. The law of iterated expectations for $n \in \mathbb{Z}_+$ now follows from *Step 1.* $\square$

C Empirics

C.1 Inflation Forecasts and Comparative Statics

Figure C.1: Inflation Forecast Accuracy Relative to Time Series Models



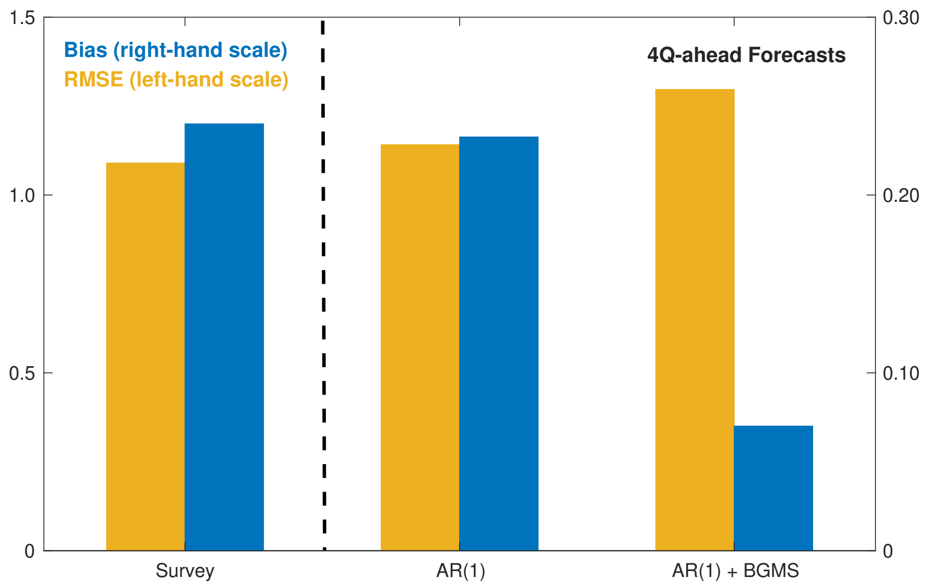
Note: Panel a illustrates the average *relative root mean-squared error* of one-quarter and four-quarter ahead forecasts of year-over-year CPI inflation from the US *Survey of Professional Forecasters* (S) relative to four time series models: *AR1* denotes forecasts from an AR(1) model, *NC* forecasts from a ("no-change") Random Walk, *SW* forecasts from an optimally-chosen time-varying parameter ARIMA model, using the criteria from Stock and Watson (2008), and BIC forecast from an ARIMA model chosen to minimize the BIC information criteria associated with one-quarter ahead forecasts. The sample period is 1982Q3:2020Q1. A RRMSE ratio below one indicates that SPF forecasts are more accurate. Panel b shows the decomposition of the model-implied root mean-squared errors of four-quarter ahead forecasts into the bias component (*Bias*) and the standard deviation of forecast errors (*Std.*) [see 2.3], and compares them to the survey data (*Survey*).

Table C.1: Over- and Underreactions of Optimal Expectations and Inflation Data

	Data Moments			*Model Moments*		
	(1)	(2)	(3)	(1)	(2)	(3)
Constant	0.24***	–	–	0.08	–	–
	(0.09)			(–)		
Average Revision	–	1.04***	–	–	1.04	–
		(0.22)			(–)	
Individual Revision	–	–	-0.20**	–	–	-0.36
			(0.07)			(–)
Observations	150	5,470	5,480	.	.	.

Note: The left-hand panel shows estimates from a regression of $y_{t+k} - f_{it}y_{t+k}$ on respondents' average- and individual forecast revisions ($I^{-1}\sum_i [f_{it}y_{t+k} - f_{it-1}y_{t+k}]$ and $f_{it}y_{t+k} - f_{it-1}y_{t+k}$), respectively. Columns (2) and (3) include individual (respondent) fixed effects and we set $k = 4$. Double-clustered robust standard errors in parentheses. Sample: 1982:Q3–2020:Q1. The right-hand panel shows model-implied moments. The estimates in column (2) and (3) are identical to those in Kohlhas and Walther (2021) Table C.7 and Broer and Kohlhas (2022) Table I. The accuracy of the model-implied forecasts and their bias can be seen from the right-two columns in Figure C.2. $*$ p<.1, $**$ p<.05, $***$ p<.01.

Figure C.2: Optimal Expectations based on an AR(1) Process for Inflation



Note: The left-two columns show the root mean-squared error and bias component of four-quarter ahead forecasts of year-over-year CPI inflation from the *US Survey of Professional Forecasters* (Survey). The sample period is 1982Q3:2020Q1. The middle-two columns (AR(1)) demonstrate the model-implied root mean-squared error and bias component of four-quarter ahead optimal expectations of output growth, assuming that y_t follows an AR(1) with persistence $\rho = 0.90$, standard deviation of innovations $\sigma_\eta = 0.60$, and mean $\mu = 2.0$. We further assume that respondents base their forecasts on $T = 14$ observations of inflation. The parameters for the AR(1) process for inflation are taken from Bordalo et al. (2020). The right-two columns (AR(1)+BGMS) show the model-implied root mean-squared error and bias component of optimal expectations, assuming that respondents observe a noisy signal of output growth, $x_{it} = y_t + \epsilon_{it}$, $\epsilon_{it} \sim \mathcal{N}\left(0, \sigma^2\right)$, where $\sigma_x = 1.37$, in each period instead of a perfect signal. We calibrate σ_x using the same approach as in Section 4.2.

Table C.2: Comparative Statics for Output Growth

Panel a: GARCH(1,1) for Output Shocks

	Estimated Process	
	Past Error	Past Vol.
GARCH(1,1)	0.48***	0.53***
	(0.15)	(0.12)
Observations	188.0	
LogLikelihood	-242.8	

Panel b: Accuracy and Shock Volatility

	Rolling Mean-squared Error	
	(1)	(2)
Garch(1,1) Volatility	1.83***	1.76***
	(0.12)	(0.13)
Time control	X	✓
Observations	188	188
F	465.5	234.1
R^2	0.72	0.72

Note: Panel a estimates a GARCH (1,1) process for year-over-year output growth, assuming that the mean of output growth follows an AR(1). Panel b shows the estimates from a linear regression of the $h-$quarter rolling mean-squared error of one-year ahead SPF forecasts of output growth on the estimated GARCH(1,1)-volatility. We set $h = 15$. The sample is: 1970:Q4-2020:Q1. The second column in Panel b controls for a linear time-trend. Robust standard errors in parentheses. $*$ p<.1, $**$ p<.05, $***$ p<.01.

C.2 Details of Time-series Models

Time-series Models: The time-series models used in Figure 2 and Figure C.1 are as follows:

- *Output*: BIC selects an ARMA(1,3), SW is an ARMA(1,1) with time-varying parameters as in Stock and Watson (2008), and NC corresponds to a Random Walk model. All models are estimated on the full sample of observations.

- *Inflation:* BIC selects an ARMA(1,3), SW is an ARMA(1,1) with time-varying parameters as in Stock and Watson (2008), and NC corresponds to a Random Walk model. All models are estimated on the full sample of observations.

Note also that we in Figure 2, because of the serial dependence of y_t, require a small change in timing for (3.3) and (3.4) to be applicable to the AR(1) case. Specifically, we in this case assume that the agent forms an estimate at time t based on y^t. The agent then observes y_{t+k}, where k is large, and forms the forecast $f_{t+k}y_{t+1+k} = \hat{\beta}\left(y^t\right)y_{t+k}$. This "behind the veil of ignorance" assumption ensures that $\mathcal{U} = -\frac{1}{2}\sigma^2 - \frac{1}{2}mse\left[\hat{\beta}\left(y^t\right)\right]\frac{\sigma^2}{1-\beta^2}$, so that the minimization of the mean-squared error of $\hat{\beta}$ is once more equivalent to the minimization of the mse. of $f_t y_{t+k}$.

Noisy Information AR(1) Model: Agent $i \in [0,1]$ has the prior belief about y_t,

$$y_t \sim \mathcal{N}\left(\mathbb{E}_{it-1}\left[y_t\right], p\right),$$

where $\mathbb{E}_{it-1}\left[y_t\right] \equiv \mathbb{E}\left[y_t \mid x_{i2}^{t-1}, \omega\right]$ evolves across time. At time t, the agent also observes

$$x_{i2t} = y_t + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{WN}\left(0, \sigma_{x2}^2\right),$$

where $\mathbb{E}\left[\epsilon_{it}\epsilon_{js}\right] = 0$ for all $j \neq i$ and $s \neq t$. We note that $\omega = \{p, \sigma_{x2}, \rho\}$.

The time t expectation of y_t conditional on agent i's information satisfies the relationship:

$$\mathbb{E}_{it}\left[y_t\right] = \gamma_1 \mathbb{E}_{it-1}\left[y_t\right] + \gamma_2 x_{i2t}, \quad \gamma_1 = 1 - \gamma_2 = \frac{p^{-2}}{p^{-2} + \sigma_{x2}^{-2}}.$$

It follows that the time t expectation of y_{t+k} equals

$$\mathbb{E}_{it}\left[y_{t+k}\right] = \beta_0 + \beta_1 x_{i1t} + \beta_2 x_{i2t} = \mathbb{E}\left[y_{t+k} \mid x_t, \beta\right],$$

where $\beta_0 = \sum_{j=0}^{k-1} \rho^j$, $\beta_1 = \rho^k \gamma_1$, and $\beta_2 = \rho^k \gamma_2$. We therefore find that

$$\begin{aligned} y_{t+k} &= \mathbb{E}_{it}\left[y_{t+k}\right] + \left(y_{t+k} - \mathbb{E}_{it}\left[y_{t+k}\right]\right) \\ &= \beta_0 + \beta_1 x_{i1t} + \beta_2 x_{i2t} + \eta_{it+k}, \quad \eta_{it+k} \sim \mathcal{N}\left(0, \sigma^2\right), \end{aligned} \tag{A31}$$

where σ^2 can be written as a function of $\sigma_{x1}^2 \equiv \rho^{2k} p + \sum_{j=1}^{k-1} \rho^{2j} \sigma_\eta^2$ and σ_{x2}^2.

C.3 Over- and Underreactions

Proof of Proposition 7: Because of the fixed shrinkage, we have that

$$f_{it}^\star y_{t+k} = m_{[k]}^\star \mathbb{E}_{it} y_{t+k}.$$

As a result, we have that individual errors and revisions equal

$$e_{it} \equiv y_{t+k} - f_{it}^\star y_{t+k} = \left(1 - m_{[k]}^\star\right) \mathbb{E}_{it} y_{t+k} + \left(y_{t+k} - \mathbb{E}_{it} y_{t+k}\right) \tag{A32}$$

$$r_{it} \equiv f_{it}^\star y_{t+k} - f_{it-1}^\star y_{t+k} = m_{[k]}^\star \mathbb{E}_{it} y_{t+k} - m_{[k+1]}^\star \mathbb{E}_{it-1} y_{t+k}. \tag{A33}$$

Case i): $\sigma_{x2}^{-2} \to \infty$. We have that $\beta_1 \to 0$ and $\beta_2 \to \rho^k$, and hence that $f_t^\star y_{t+k} = m_{[k]}^\star \mathbb{E}_t y_{t+k} = m_{[k]}^\star \left(\beta_0 + \rho^k y_t\right)$. Using (A32) and (A33), we therefore find that

$$\mathbb{C}ov\left(e_{it}, r_{it}\right) = (1 - m_{[k]}^\star) m_{[k]}^\star \rho^{2k} \left(1 - \frac{m_{[k+1]}^\star}{m_{[k]}^\star} \rho^2\right) \mathbb{V}\left[y_t\right],$$

which is positive if $m^\star_{[k+1]} \in \left(m^\star_{[k]}, 1/\rho^2 m^\star_{[k]}\right)$.

Case ii): $\sigma_{x2}^{-2} \to 0$. We have that $\beta_1 \to \rho^k$ and $\beta_2 \to 0$, and hence that $f_t^\star y_{t+k} = m^\star_{[k]} \mathbb{E}_t y_{t+k} = m^\star_{[k]} \left(\beta_0 + \rho^k x_{i1t}\right)$. Using (A32) and (A33), we therefore find

$$\mathbb{C}ov\left(e_{it}, r_{it}\right) = \left(1 - m^\star_{[k]}\right) \left(m^\star_{[k]} - m^\star_{[k+1]}\right) \mathbb{V}\left[\mathbb{E}_{it-1} y_{t+k}\right],$$

where $\mathbb{E}_{it-1} y_{t+k} = \beta_0 + \rho^k x_{i1t}$. We note that $\mathbb{C}ov\left(e_{it}, r_{it}\right)$ is negative if $m^\star_{[k+1]} > m^\star_{[k]}$.

The proof is then completed by the fact that the slope coefficient in the regression studied in the proposition is the (cross-sectional) average of the slope coefficients in the individual regressions of e_{it} on r_{it}. The signs of the latter coefficients, in turn, depend only on *sign* $\left\{\mathbb{C}ov\left(e_{it}, r_{it}\right)\right\}$. $\square$

D Applications

D.1 Properties of Optimal Expectations

Lemma D.1. *Suppose an agent wishes to minimize her expectation $f_t^\star[\cdot]$ of*

$$\mathcal{U} = Q\left(c_t, y_{t+k}\right), \tag{A34}$$

where the function $Q : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ is a quadratic polynomial whose coefficients are in Ω_t, c_t represents the agent's choice variable, and $y_{t+k} \in \mathbb{R}^p$ is a vector of (potentially random) variables. Then,

$$\frac{\partial}{\partial c_t} f_t^\star\left[Q\left(c_t, y_{t+k}\right)\right] = f_t^\star\left[\frac{\partial}{\partial c_t} Q\left(c_t, y_{t+k}\right)\right].$$

Proof of Lemma D.1: We have that

$$\begin{aligned}
\frac{\partial}{\partial c_t} f_t^\star\left[Q\left(c_t, y_{t+k}\right)\right] &= \lim_{h\to 0} \frac{1}{h}\left\{f_t^\star\left[Q\left(c_t+h, y_{t+k}\right)\right] - f_t^\star\left[Q\left(c_t, y_{t+k}\right)\right]\right\} \\
&= \lim_{h\to 0} f_t^\star\left\{\frac{Q\left(c_t+h, y_{t+k}\right) - Q\left(c_t, y_{t+k}\right)}{h}\right\} \\
&= \lim_{h\to 0} f_t^\star\left\{\frac{\partial}{\partial c_t} Q\left(\bar{c}(h), y_{t+k}\right)\right\},
\end{aligned}$$

where $\bar{c}(h) \in (c_t, c_t+h)$ exists by the mean value theorem. The second line follows from the linearity of the expectation operator. Furthermore, because Q is a quadratic polynomial:

$$\lim_{h\to 0} f_t^\star\left\{\frac{\partial}{\partial c_t} Q\left(\bar{c}(h), y_{t+k}\right)\right\} = \lim_{h\to 0} f_t^\star\left\{\alpha_0 + \alpha_1\bar{c}(h) + \alpha_2' y_{t+k}\right\},$$

where $\alpha_0, \alpha_1, \alpha_2 \in \Omega_t$ and $\bar{c}(h) \in \Omega_t$. Thus,

$$\begin{aligned}
\lim_{h\to 0} f_t^\star\left\{\frac{\partial}{\partial c_t} Q\left(\bar{c}(h), y_{t+k}\right)\right\} &= \alpha_0 + \alpha_1 \lim_{h\to 0} \bar{c}(h) + \alpha_2' f_t^\star y_{t+k} \\
&= \alpha_0 + \alpha_1 c_t + \alpha_2' f_t^\star y_{t+k} = f_t^\star\left[\frac{\partial}{\partial c_t} Q\left(c_t, y_{t+k}\right)\right].
\end{aligned}$$

This completes the proof. □

D.2 Consumption and Expectation Choices

Stage 2: We start by solving the agent's consumption-choice problem at $t \geq 1$, assuming her optimal expectations satisfy the Law of Iterated Expectations. The agent's problem is

$$\max_{\{c_{t+k}\}_k} f_t\left[\sum_{k=0}^{T-t} \delta^{k-1} u\left(c_{t+k}\right)\right] \quad \text{s.t.} \quad \sum_{k=0}^{T-t} R^{-k}\left(c_{t+k} - y_{t+k}\right) = a_t, \tag{A35}$$

where $a_1 \geq 0$. Lemma 1 implies that we can use standard steps to show that

$$c_t = f_t c_{t+1}. \tag{A36}$$

Consider now the budget constraint $\sum_{k=0}^{T-t} R^{-k}(c_{t+k} - y_{t+k}) = a_t$. Recursively taking expectations $f_{T-1}[\cdot]$, $f_{T-2}[\cdot]$,..., $f_t[\cdot]$ of this expression, using the LIE and (A36), shows that

$$c_t \sum_{k=0}^{T-t} R^{-k} = a_t + y_t + \sum_{k=1}^{T-t} R^{-k} f_t y_{t+k},$$

Hence, we conclude that

$$c_t = \theta_t \left(Ra_t + y_t + \sum_{k=1}^{T-t} R^{-k} f_t y_{t+k} \right) = \theta_t \left(Ra_t + y_t + \sum_{k=1}^{T-t} \lambda^{k-1} R^{-k} f_t y_{t+1} \right), \tag{A37}$$

where $\theta_t \equiv \frac{1-R^{-1}}{1-R^{-(T-t+1)}}$ and the second equality uses that $f_t y_{t+k} = \lambda^{k-1} f_t y_{t+1}$.

The agent's assets are $t \geq 2$ equal $a_t = Ra_{t-1} + y_{t-1} - c_{t-1}$.

Stage 1: We next turn to the agent's expectation-formation problem. We start with a convenient re-statement of the agent's consumption-choice problem in the second stage. Let $\mathcal{L}(\mathbf{c})$ denote the Lagrangian associated with (A35) in the initial period, and denote the corresponding perfect-foresight version with $\mathcal{L}^\star(\mathbf{c})$. Thus, $\mathcal{L}(\mathbf{c}) = f_1 \mathcal{L}^\star(\mathbf{c})$. Finally, let $\mathbf{c}^\star$ denote the solution to the perfect-foresight problem.

Notice that because the utility function $u(\cdot)$ is assumed quadratic, we have that

$$\begin{aligned}
\mathcal{L}^\star(\mathbf{c}) &= \mathcal{L}^\star(\mathbf{c}^\star) + \frac{\partial \mathcal{L}^{\star\prime}}{\partial \mathbf{c}}_{|\mathbf{c}=\mathbf{c}^\star} (\mathbf{c} - \mathbf{c}^\star) + \frac{1}{2} (\mathbf{c} - \mathbf{c}^\star)' \frac{\partial^2 \mathcal{L}^\star}{\partial \mathbf{c}^2}_{|\mathbf{c}=\mathbf{c}^\star} (\mathbf{c} - \mathbf{c}^\star) \\
&= \mathcal{L}^\star(\mathbf{c}^\star) + \frac{1}{2} (\mathbf{c} - \mathbf{c}^\star)' \frac{\partial^2 \mathcal{L}^\star}{\partial \mathbf{c}^2}_{|\mathbf{c}=\mathbf{c}^\star} (\mathbf{c} - \mathbf{c}^\star) \\
&= \mathcal{L}^\star(\mathbf{c}^\star) + \frac{1}{2} u'' \sum_t \delta^{t-1} (c_t - c_t^\star)^2,
\end{aligned} \tag{A38}$$

where $u'' < 0$ and constant. We can therefore re-cast the agent's consumption problem in the initial period as choosing $\{c_t\}$ to minimize the agent's expected value of $\sum_t \delta^{t-1}(c_t - c_t^\star)^2$.

We can now use this re-cast consumption-problem to solve for the agent's optimal expectations. It follows from (A37) and (A38) that the agent will choose $\{f_t[\cdot]\}$ to minimize

$$\begin{aligned}
\mathbb{E} \sum_t \delta^{t-1} (c_t - c_t^\star)^2 &= \mathbb{E} \sum_t \delta^{t-1} \eta_t^2 \left[\sum_{k=1}^{T-t} R^{-k} (f_t y_{t+k} - y_{t+k}) \right]^2 \\
&= \mathbb{E} \sum_t \delta^{t-1} \eta_t^2 \left[\sum_{k=1}^{T-t} R^{-k} \lambda^{k-1} (f_t y_{t+1} - y_{t+1}) + \text{shocks}_{t+1:T} \right]^2,
\end{aligned}$$

where $\text{shocks}_{t+1:T}$ denote terms related to shocks from $t+1$ onwards, uncorrelated with period-t information. Thus, the agent will choose $f_t[\cdot]$ to minimize $\sum_t \delta^{t-1} \eta_t^2 \left(\sum_{k=1}^{T-t} R^{-k} \lambda^{k-1} \right)^2 \mathbb{E}\left[(f_t y_{t+1} - y_{t+1})^2 \right]$. That is, the agent will choose $f_t[\cdot]$ to minimize the one-period ahead *mse* of her income forecast,

$$\mathbb{E}\left[(y_{t+1} - f_t y_{t+1})^2 \right].$$

We conclude that $\{f_t y_{t+1}\} = \{f_t^\star y_{t+1}\}$, and that the agent's expectations satisfy the LIE.

D.3 Consumption Implications

Proof of Corollary 1: The myopia of consumption choices and the predictability of consumption changes follow from discussions in the main text. The average level of consumption for $t = 1, 2, .., \tau, ...T$ conditional on $x_{1t} = \mu_x > 0$ and $a_1 \geq 0$ is, using (A37) and the equation of motion for a_t:[5]

$t = 1:$

$$\mu_{c,1} \equiv \mathbb{E}\left[c_1 \mid x_1^1 = \mu_x\right] = \theta_1 \left(Ra_1 + \beta_1\mu_x + \sum_{k=1}^{T-1} R^{-k} \times m_1^\star \beta_1 \mu_x\right)$$

$$= \frac{1-R^{-1}}{1-R^{-T}}\left[Ra_1 + \left(1 + \frac{R^{-1}-R^{-T}}{1-R^{-1}} m_1^\star\right)\beta_1\mu_x\right] \tag{A39}$$

$$\mu_{a,2} \equiv \mathbb{E}\left[a_2 \mid x_1^1 = \mu_x\right] = Ra_1 + \beta_1\mu_x - \mu_{c,1} \tag{A40}$$

$t = 2:$

$$\mu_{c,2} = \frac{1-R^{-1}}{1-R^{-(T-1)}}\left[R\mu_{a,2} + \left(1 + \frac{R^{-1}-R^{-(T-1)}}{1-R^{-1}} m_2^\star\right)\beta_1\mu_x\right] \tag{A41}$$

$$\mu_{a,3} = R\mu_{a,2} + \beta_1\mu_x - \mu_{c,2} \tag{A42}$$

$\vdots$

$t = \tau:$

$$\mu_{c,\tau} = \frac{1-R^{-1}}{1-R^{-(T-t+1)}}\left[R\mu_{a,\tau} + \left(1 + \frac{R^{-1}-R^{-(T-\tau+1)}}{1-R^{-1}} m_\tau^\star\right)\beta_1\mu_x\right] \tag{A43}$$

$$\mu_{a,\tau+1} = R\mu_{a,\tau} + \beta_1\mu_x - \mu_{c,\tau}. \tag{A44}$$

Notice that if $m_t^\star = 1$, as in the fully-informed case, then $\mu_{c,t}$ in (A39) to (A43) is constant across time. By contrast, because of the uncertainty about the best use of information, in our case, $m_t^\star \in (0, 1)$ and increasing in t (Section 3.4). This makes $\mu_{t,c}$ in (A43) smaller than its informed value for low t and increasing with time. The latter in part also due to assets in (A44) initially being higher than in the informed case. Finally, notice that the average expectation in the noisy-information case, $\mathbb{E}\left\{\mathbb{E}\left[y_{t+1} \mid \Omega_t\right] \mid x_1^t = \mu_x\right\}$, where $\Omega_t = \left\{y^t, \beta_1\right\}$, is equal to the fully-informed value $\beta_1\mu_x$. This follows from the LIE. □

Proof of Corollary 2: Consider the consumption response to a productivity shock at date t, and imagine that we increase x_{1t} by $\epsilon > 0$. All other shocks are kept at their mean values, and

[5] Recall that we assume that the household observes (x_0, y_1) and (x_{-1}, y_0) in the initial period, so that the household can construct an estimate of β_1 in period $t = 1$.

we fix $\hat{\beta}_1^\star = m_t^\star \beta_1$. From equation (A37), we have for $\tau \geq t$:

$$\begin{aligned} c_\tau &= \theta_\tau \left(Ra_\tau + \beta_1 \times \lambda^{\tau-t}\epsilon + \gamma_\tau \times \beta_1 m_t^\star \lambda^{\tau-t+1}\epsilon \right) \\ a_{\tau+1} &= Ra_\tau + \beta_1 \times \lambda^{\tau-t}\epsilon - c_\tau, \end{aligned}$$

where $\gamma_\tau \equiv \sum_{k=1}^{T-\tau} \lambda^{k-1} R^{-k}$ and $a_t = a_1 \geq 0$. Notice that the consumption response is initially smaller than under full-information, where $m_t^\star = 1$. It is also more persistent, due to the larger initial accumulation of assets $a_{\tau+1}$ with $R \geq 1$, which feed into future consumption $c_{\tau+k}$. The final result in the statement follows from the increase of $m_t^\star$ in x_{1t}^2 in (3.8). $\square$

Proof of Corollary 3: Consider the expected value of c_t in (A37) when $x_{t-1} = x_t = \mu_x > 0$, $a_t = a_1 \geq 0$, and we fix $\hat{\beta}_1^\star = m_t^\star \beta_1$. The result then follows from the decrease of $m_t^\star$ in σ in (3.8). $\square$

D.4 Other Models of Income Expectations

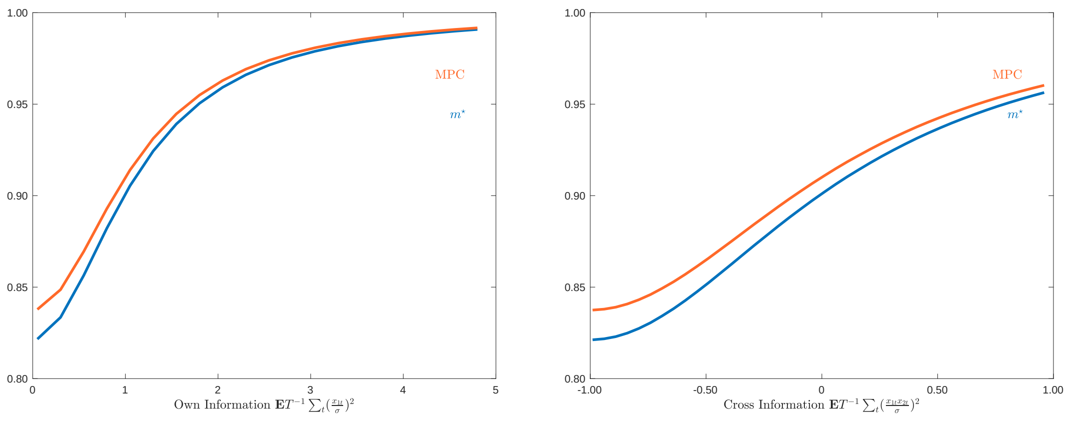
Suppose household income at time $t + k$ follows:

$$y_{t+k} = \beta_1 x_{1t} + \beta_2 x_{2t} + \eta_{t+k}, \quad \eta_{t+k} \sim \mathcal{N}(0, 1), \tag{A45}$$

where $x_{1t} \sim \mathcal{N}\left(0, \sigma_1^2\right)$ and $x_{2t} = \sigma_{12} x_{1t} + u_{2t}$ with $u_{2t} \sim \mathcal{N}\left(0, \sigma_2^2\right)$ independent of x_{1t}.

Figure D.1 shows the effects of expected own- and cross-information on a household's caution choice $m^\star$ and marginal propensity to consume out of permanent income.

Figure D.1: Comparative Statics of Own- and Cross-information Effects



Note: The figure illustrates the effects of expected own- and cross- information on a household's caution choice and marginal propensity to consume (MPC) out of permanent income (Section 5.2). Throughout, we set $T = 20$, $t = 13$ $\sigma = \sigma_2 = 1$, $\beta_1 = \beta_2 = 0.50$, and let $R = 1$. The left-hand panel varies σ_1 to attain the desired own-information term, letting $\sigma_{12} = 0$. The right-hand panel, by contrast, varies σ_{12} to achieve the desired cross-information term, letting $\sigma_1 = \sigma_2$.

D.5 Optimal Sticky Prices and Monetary Policy

We proceed with a model of monetary policy in the spirit of Maćkowiak and Wiederholt (2009). The economy is comprised of a continuum of measure one of firms, indexed by $i \in [0, 1]$. At the start of every period $t = 1, 2, \ldots$ each firm sets the price of its own good P_{it}, so as to maximize its optimal expectation of the discounted sum of profits

$$\sum_{\tau=t}^{\infty} \delta^{\tau-t} \pi \left(P_{i\tau}, P_\tau, Q_\tau, Z_\tau\right), \tag{A46}$$

where $\delta \in (0, 1)$, $P_t \equiv \exp \int_0^1 \log P_{it} di$ denotes the economy-wide price index, Q_t aggregate output, and Z_t firm productivity. We assume that the profit function π is twice continuously differentiable and homogenous of degree zero in its first two arguments. We further assume that π is single-peaked in P_{it}. These assumptions are, for example, satisfied by the standard model of monopolistic competition. Firm productivity evolves in accordance with $\log\left(Z_t\right) \sim \mathcal{N}\left(\mu_z, \sigma_z^2\right)$ and is known to firms at the start of each period. Turning to the demand side of the economy, the central bank sets nominal demand $\mathcal{M}_t = P_t Q_t$ so that

$$\log\left(\mathcal{M}_t\right) = \beta_1 \log\left(\mathcal{M}_{t-1}\right) + \eta_t, \quad \eta_t \sim \mathcal{N}\left(0, \sigma^2\right), \tag{A47}$$

where $\beta_1 \in (0, 1)$, $\mathbb{E}\left[\eta_t \eta_s\right] = 0$ for all $t \neq s$, and $\mathbb{E}\left[\eta_t \log Z_s\right] = 0$ for all t and s. Nominal demand is determined after firms set prices for the period. Firms' information set at time t is therefore $\Omega_t = \left\{\log \mathcal{M}^{t-1}, \log Z^t; \pi(\cdot), \sigma\right\}$. We note that firms do not know the true value of β_1. Finally, to close the model, we assume that firms choose how to form expectations *ex ante,* before the realization of any shocks, consistent with (1). Firms then in each period form optimal expectations and set prices based on the information that realizes.

A few simple derivations show that this economy is nested within our framework. Let v_t denote the log-deviation of the random variable V_t from its non-stochastic value. Identical steps to those in Maćkowiak and Wiederholt (2009) then show that a second-order approximation of a firm's profit function allows one to re-state a firm's pricing problem as maximizing $-\frac{1}{2} f_t^\star (p_{it} - p_t^\star)^2$ with respect to p_{it} (see the proof of Corollary D.1), where $p_t^\star = p_{it}^\star$ denotes the flexible FIRE price:

$$p_t^\star = \mu_t + \frac{\pi_{14}}{\pi_{13}} z_t, \quad \mu_t \equiv \log\left(\mathcal{M}_t\right). \tag{A48}$$

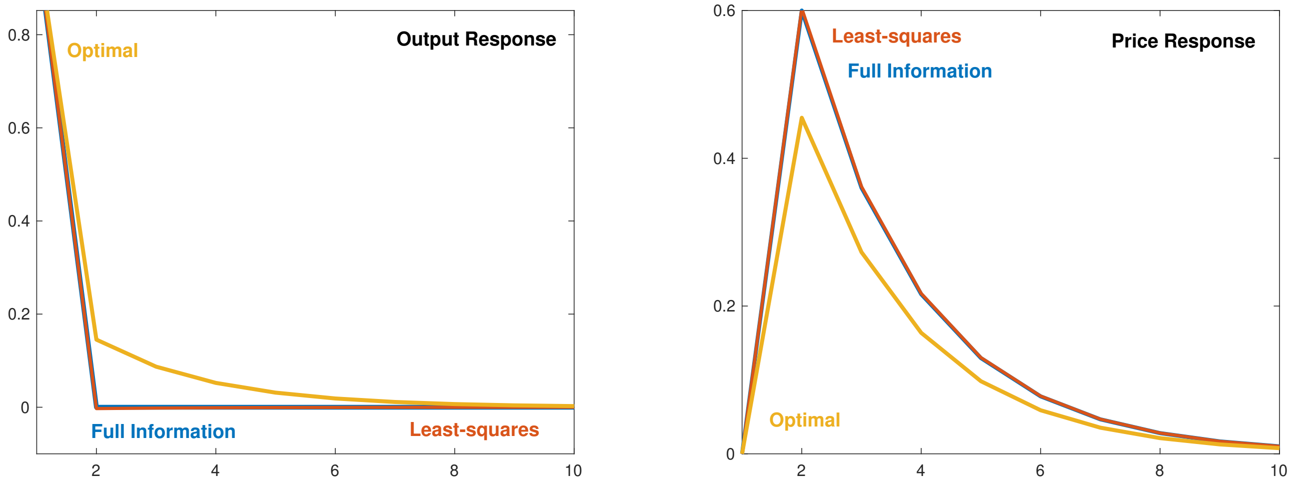
A firm's optimal price p_{it} therefore equals its optimal expectation of $p_t^\star$:

$$p_{it} = f_t^\star \left[p_t^\star\right] = f_t^\star \left[\mu_t \mid \mu^{t-1}\right] + \frac{\pi_{14}}{\pi_{13}} z_t = \hat{\beta}_1^\star \mu_{t-1} + \frac{\pi_{14}}{\pi_{13}} z_t \quad \forall i, \tag{A49}$$

where $\hat{\beta}_1^\star$ is given by Proposition 3.[6] Lastly, because second-order profits depend only on the squared distance between a firm's price and its FIRE value, we can write a firm's ex ante

[6] Notice that we deal with the AR(1) structure of (A47) in the same manner as in Section 4.

Figure D.2: Impulse Response Functions to a Monetary Policy Shock



The figure depicts the median impulse response of output q_t (left panel) and the price level p_t (right panel) in response to a one standard deviation increase in nominal demand μ_t. The yellow lines indicate responses for an economy in which firms form optimal expectations; the blue lines an economy in which firms form FIRE expectations; and lastly the red lines an economy in which firms use (adj.) least-squares to construct their expectations. Because the latter is unbiased, the red and blue lines are (close to) identical. To parameterize the economy, we set $\frac{\pi_{14}}{|\pi_{13}} = -0.15$, $\beta = 0.60$, $\sigma = 1$, and $T = 10$. The parameters that determine the curvature of firms' profit function are taken from Maćkowiak and Wiederholt (2009). The number of observations T corresponds to double the average age of a firm in the latest round of the BDS survey.

expectation problem as maximizing $-\frac{1}{2}\mathbb{E}(p_t^\star - f_t p_t^\star)^2 = -\frac{1}{2}\mathbb{E}(\mu_t - f_t\mu_t)^2$ with respect to $f_t\mu_t$. We conclude that a firm's problem is nested within our framework with $y_{t+k} = \mu_t$ and $x_{1t} = \mu_{t-1}$. The following is then a consequence of Proposition 3 and 5:

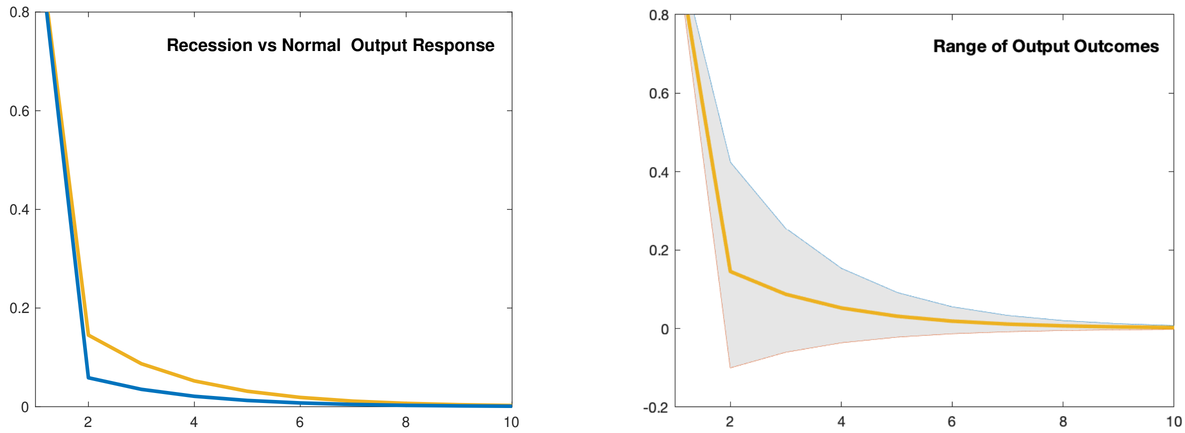
Corollary D.1. *In response to a monetary policy shock η_t, the price level p_{t+k}, on average, adjusts by less than one-for-one, and monetary policy has real effects on output q_{t+k} for $k \geq 0$. Furthermore, monetary policy has smaller effects after large movements in μ_t.*

That monetary policy has real effects at time t is just a consequence of the assumption of one-period ahead preset prices. What is interesting is that the real effects of monetary policy persist for future periods $t + k$. This is because firms optimally down-weigh their responses to new information about nominal demand μ_t. Prices, on average, do not fully adjust in response to changes in nominal demand even after firms are allowed to reset their prices. Monetary policy as a consequence has protracted real effects (Figure D.2).

The results in Corollary D.1 closely resemble those in Lucas (1972), Woodford (2002), Nimark (2008), Maćkowiak and Wiederholt (2009), and Angeletos *et al.* (2016), among others, where monetary policy has protracted real effects because firms observe noisy information. Similar to such models, monetary policy has real effects because firms do not fully update their expectations and hence prices. However, Corollary D.1 and (A49) also demonstrate important differences from models in which noisy information is the underlying cause of price stickiness.

First, unlike in, for example, Woodford (2002), the real effects of monetary policy are history-

Figure D.3: History-dependent Responses to a Monetary Policy Shock



The figure depicts the median impulse response of output q_t (left panel) in response to a one standard deviation increase in nominal demand μ_t when firms form optimal expectations in accordance with (A49). We depict this impulse response in two cases: (i) in which the economy only experiences average ("normal") realizations of μ_t; and (ii) in which the economy always experiences a one standard deviation decline in μ_t in the final two periods before the increase in nominal demand ("recession"). The right panel depicts the 1 and 99 percentile realizations of the impulse response function in the "normal" case across different simulations of the history of observations before the monetary impulse. All parameters are set to the same values as those in Figure D.2.

dependent. The precise realization of the fundamental μ_t matters for firms' implied caution choices, and hence for how much firms choose to down-weigh their responses on average (Proposition 3; equation A49). Figure D.3 depicts the range of outcomes that can arise after an increase in nominal demand for a simple calibration of the model. Notice that, while monetary policy has positive real effects on average, the range of outcomes is sizable. Further, similar to the empirical results in Lucas (1973), monetary policy has smaller real effects in economies that have experienced more volatile fundamentals. This is because firms' prices in these economies are more flexible; firms are on average more responsive to new information, and hence adjust their prices by more in reaction to shocks. Indeed, through the lens of our equivalence result in Proposition **6**, it as *as-if* firms pay "more attention" after large shocks.

Second, this history-dependence can also help explain facets of the data that seem at odds with simple noisy information models of price stickiness. For example, Vavra (2014) and Tenreyro and Thwaites (2016) document that monetary policy has had smaller effects in recessions over the past 50 years than expansions, because prices are here more flexible. This is inconsistent with simple noisy information models of price stickiness, where monetary policy has identical effects in normal and recession times. Combined with the skew in output growth that has existed over the past 50 years (most large changes in output have occurred during recessions; Veldkamp, 2005), the above model can however be consistent with this evidence.

Figure D.3 shows the impulse response of output to a one standard deviation increase in nominal demand when firms have lived for ten years (c. double the average age of a firm in the US BDS firm census). We contrast an economy which has only experienced median realizations of μ_t with one that in the final two periods before the increase in nominal demand experiences a one standard deviation fall in μ_t and hence output. Consistent with the findings of Vavra (2014) and Tenreyro and Thwaites (2016), we find that monetary policy, on average, has smaller effects on the economy that experiences the recession.

Proof of Corollary D.1: Let $\tilde{\pi}$ denote the second-order approximation of a firm's profit function around the origin,

$$\tilde{\pi}\left(p_{it}, p_t, q_t, z_t\right) = \pi_1 p_{it} + \frac{\pi_{11}}{2} p_{it}^2 + \pi_{12} p_{it} p_t + \pi_{13} p_{it} y_t + \pi_{14} p_{it} z_t + t.u.p., \tag{A50}$$

where $t.u.p.$ denotes *terms unrelated to* p_{it}.

It follows that a firm's full information (FIRE) flex-price choice satisfies

$$p_{it}^{\star} = p_t^{\star} + \frac{\pi_{13}}{\mid \pi_{11} \mid} q_t + \frac{\pi_{14}}{\mid \pi_{11} \mid} z_t = \left(1 - \frac{\pi_{13}}{\mid \pi_{11} \mid}\right) p_t^{\star} + \frac{\pi_{13}}{\mid \pi_{11} \mid} \mu_t + \frac{\pi_{14}}{\mid \pi_{11} \mid} z_t,$$

so that in a symmetric equilibrium

$$p_t^{\star} = \mu_t + \frac{\pi_{14}}{\pi_{13}} z_t. \tag{A51}$$

We can furthermore use (A50) to derive an approximation of the difference between a firm's valuation of its profits π_{it} and those that would have arisen under full information and flexible prices $\pi_{it}^{\star}$:

$$\begin{aligned} \tilde{\pi}_{it} - \tilde{\pi}_{it}^{\star} &= \frac{\pi_{11}}{2} p_{it}^2 - \frac{\pi_{11}}{2} \left(p_{it}^{\star}\right)^2 + \left(\pi_{12} p_t + \pi_{13} y_t + \pi_{14} z_t\right)_{|\star} \left(p_{it} - p_t^{\star}\right) \\ &= \frac{\pi_{11}}{2} p_{it}^2 - \frac{\pi_{11}}{2} \left(p_{it}^{\star}\right)^2 - \left(\pi_{11} p_t^{\star}\right) \left(p_{it} - p_t^{\star}\right) = \frac{\pi_{11}}{2} \left(p_{it} - p_t^{\star}\right)^2, \end{aligned} \tag{A52}$$

where the second equality exploits the first-order condition to the firm's profit maximization problem in (A50) under full information and flexible prices. Finally, it follows from (A52) that a firm's optimal price equals

$$p_{it} = p_t = f_t^{\star}\left[p_t^{\star}\right], \tag{A53}$$

where (A52) allows us to characterize a firm's second-stage problem $\min_{p_{it}} f_t^{\star}\left(p_{it} - p_t^{\star}\right)^2$. The statement now follows from an application of Proposition 3 and 5 to (A53) and (A51). □

References

ANGELETOS, G.-M., LA'O, J. and IOVINO, L. (2016). Real rigidity, nominal rigidity, and the social value of information. *American Economic Review*, **106(1)**, 200–227.

BROCKWELL, P. and DAVIS, R. A. E. (1991). *Time series: theory and methods: theory and methods*. Springer Science & Business Media.

LUCAS, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review*, **63** (3), 326–334.

LUCAS, R. E. J. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, **4** (2), 103–124.

MAĆKOWIAK, B. and WIEDERHOLT, M. (2009). Optimal sticky prices under rational inattention. *The American Economic Review*, **99** (3), 769–803.

NIMARK, K. (2008). Dynamic pricing and imperfect common knowledge. *Journal of Monetary Economics*, **55** (2), 365–382.

STOCK, J. H. and WATSON, M. W. (2008). Phillips curve inflation forecasts.

TENREYRO, S. and THWAITES, G. (2016). Pushing on a string: Us monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics*, **8** (4), 43–74.

VAVRA, J. (2014). Inflation dynamics and time-varying volatility: New evidence and an ss interpretation. *The Quarterly Journal of Economics*, **129** (1), 215–258.

VELDKAMP, L. L. (2005). Slow boom, sudden crash. *Journal of Economic Theory*, **124** (2), 230–257.

WOODFORD, M. (2002). *Imperfect Common Knowledge and the Effects of Monetary Policy*. Nber working papers, Department of Economics, Columbia University.