

Measuring Biases in Expectation Formation^{*}

Simas Kučinskas[†]

Florian S. Peters[‡]

Abstract

We develop a general framework for measuring biases in expectation formation. The basic insight is that biases can be inferred from the impulse response function of forecast errors. The method does not require knowing the true data-generating process and is straightforward to apply empirically. Our theoretical framework encompasses all major models of expectations, and it yields a set of new empirical predictions. In an application on inflation expectations, we (i) find underreaction in both individual and consensus forecasts; and (ii) use the new empirical predictions to rank the performance of different models.

Keywords: expectation formation; bias; underreaction; overreaction.

JEL Classification: C53, D83, D84, E70, G40.

^{*}For comments and suggestions at various stages of the project, we thank Frank de Jong, Andreas Fuster, Hans-Martin von Gaudecker, Yueran Ma, Augustin Landier, Ioanid Rosu, Christian Stoltenberg, Paul Tetlock, Uwe Thümmel, and Michael Weber, as well as seminar audiences at the Bank of Lithuania, HEC Paris, University of Amsterdam, University of Bonn, and Vrije Universiteit Amsterdam. Replication files (with Python and STATA code) are available [online](#).

[†]University of Amsterdam (e-mail: s.kucinskas@uva.nl, website: simaskucinskas.com).

[‡]University of Amsterdam (e-mail: f.s.peters@uva.nl, website: florianpeters.org).

1 Introduction

There is ample evidence that subjective expectations deviate from simple forms of rational expectations.¹ However, there is little agreement on how subjective expectations are actually formed. The lack of consensus has led to a proliferation of models, some of them taking very different views on expectation formation. For instance, much research in macroeconomics has focused on models featuring *underreaction* to new information.² At the same time, many prominent models in financial economics exhibit *overreaction*.³ Even some of the empirical evidence seems conflicting, with some findings supporting underreaction, and others more consistent with overreaction.⁴

We propose a general framework for measuring biases in expectation formation. Our proposed method has several attractive features. First, the method does not require precise knowledge of the true data-generating process. Second, the framework is able to capture rich forms of bias. Finally, our method is straightforward to apply empirically. The method can be used whenever both expectations and realized values are observed, and the time series dimension of the dataset is large enough.

The basic insight is that biases can be inferred from the response of forecast errors to past news. To be concrete, consider an analyst who is forecasting future inflation. For some reason, inflation in the current quarter is higher than was expected by the analyst. If the analyst reacts to new information optimally, the forecast error in the current quarter should not be predictive of future forecast errors. However, suppose that the analyst does not follow macroeconomic developments very carefully and tends to underreact to news. Since inflation is persistent, a positive forecast error today implies that the forecast error next quarter is likely to again be positive. As a result, underreaction leads to positively autocorrelated forecast errors. In contrast, if the analyst overreacts to the higher-than-expected inflation, inflation in the next quarter will on average be lower than

¹ For surveys, see [Pesaran and Weale \(2006, Section 5\)](#) and [Manski \(2018\)](#). [Coibion, Gorodnichenko, and Kamdar \(2018\)](#) provide another recent overview, focusing on inflation expectations. We use the adjective “simple” because even if agents are rational and have full information, their forecasts may differ from the true conditional expectations (e.g., if they have asymmetric loss functions, see [Patton and Timmermann \(2007\)](#) and references therein).

² Examples include sticky information ([Mankiw and Reis, 2002](#)), rational inattention ([Sims, 2003](#)), imperfect information (e.g., [Woodford, 2003](#)), and sparsity-based models of limited attention ([Gabaix, 2017b, 2018](#)).

³ Examples include diagnostic expectations ([Gennaioli and Shleifer, 2010](#); [Bordalo, Gennaioli, and Shleifer, 2018b](#)), extrapolative expectations (e.g., [Cutler, Poterba, and Summers, 1990b](#); [DeLong, Shleifer, Summers, and Waldmann, 1990](#); [Barberis, Greenwood, Jin, and Shleifer, 2015](#)), and overconfidence (e.g., [Daniel, Hirshleifer, and Subrahmanyam, 1998](#); [Odean, 1998](#)).

⁴ A striking example is given by [De Bondt and Thaler \(1990\)](#) and [Abarbanell and Bernard \(1992\)](#). These papers provide evidence suggestive of both overreaction (De Bondt and Thaler) and underreaction (Abarbanell and Bernard) for the case of stock market analysts. See also [Bouchaud, Krüger, Landier, and Thesmar \(2018\)](#).

anticipated. With overreaction, forecast errors tend to be negatively autocorrelated.

We take this logic to its natural conclusion by representing biases in expectation formation as an *impulse response function* (IRF) of forecast errors. The IRF of forecast errors yields a set of *bias coefficients* which provide a natural measure of under- or overreaction to news at various lags. If expectations react to news in an unbiased way, a shock to forecast errors should die out immediately. If the shock does not die out immediately, expectations are biased. Since we do not impose any structure on the bias coefficients, the method is able to capture rich forms of bias. Empirically, we use local projections (Jordá, 2005) to estimate the IRF of forecast errors without imposing strong parametric assumptions.

The estimated bias coefficients could have multiple economic interpretations. Non-zero bias coefficients only provide evidence of statistical bias. For example, the inflation analyst may underreact to news about inflation because of informational frictions, psychological biases, or career concerns. Additional structure is necessary to distinguish between these competing explanations. Nevertheless, if we choose a particular model of expectations and postulate a process for the variable being forecast, it is straightforward to derive the implied bias coefficients. Comparing the estimated bias coefficients to their theoretical counterparts provides a natural test of the model. Bias coefficients also provide a natural set of moments to target for calibration exercises or structural estimation.

The existing literature on expectations is voluminous, and we refer to the surveys cited above (Footnote 1) for comprehensive reviews. We are certainly not the first to study the predictability of forecast errors. A key result in the literature on forecast evaluation is that with optimal one-step-ahead forecasts, the resulting forecast errors are white noise (see, e.g., Diebold and Lopez, 1996). Our contribution is to show that the structure of autocorrelation, not just its existence, is informative about how expectations are formed. In the paper that introduced rational expectations, Muth (1961, pp. 321–322) already considered a model of biased expectation formation that is a special case of our framework. More recently, Coibion and Gorodnichenko (2012) proposed an empirical technique that is related to our method but requires more information about the underlying process. We discuss these and other related papers in detail in Section 2.3 after presenting our framework.

Related issues have also been extensively studied outside economics, most prominently in psychology. The psychology literature has documented a number of biases in the way people form subjective beliefs. Experimental studies of Bayesian updating have found that subjects often do not update enough, a finding known as conservatism bias (Edwards, 1968). Studies on belief persistence show that people often hold on

to incorrect beliefs (see, e.g., [Nickerson, 1998](#), pp. 187–188). Conservatism bias and belief persistence are both forms of underreaction to new information. However, other well-known findings in psychology are more consistent with overreaction. For example, [Kahneman and Tversky \(1973\)](#) find that subjects fail to incorporate base rates and the reliability of the available information when making predictions. The famous hot-hand fallacy study of [Gilovich, Vallone, and Tversky \(1985\)](#) suggests that people overreact to noise.⁵

The general framework is developed in [Section 2](#). The basic setting is that of an agent forecasting some variable of interest. We consider expectations that can be represented as a linear combination of past shocks. The key result is that biases in expectation formation are given by the IRF of the forecast errors.

In [Section 3](#) we show that the framework is sufficiently flexible to nest all major existing models of expectations. As a side benefit, the IRF representation of biases provides a new lens to look at existing models. [Section 4](#) discusses a number of extensions to the basic framework, including multiple-step-ahead forecasts, heterogeneity in expectations, and measurement error. We provide a Monte Carlo study in [Section 5](#) to gauge the statistical power of our procedure and investigate its performance when some of the assumptions are violated.

In [Section 6](#) we apply the method to data on inflation forecasts from the Survey of Professional Forecasters. The results show underreaction for up to one year after the arrival of news, both when individual and consensus forecasts are used. We also discuss how the estimated bias coefficients can be used to distinguish between different models of expectations. [Section 7](#) concludes by discussing limitations of the proposed methodology and outlining potential directions for future work.

2 Methodology

2.1 Framework

We want to measure biases in how an agent forms expectations about some variable x_t . To build intuition, we first study the situation in which x_t follows a univariate process. Then, we generalize to the multivariate case. Throughout, we suppose that any

⁵ This interpretation has recently been challenged by [Miller and Sanjurjo \(2018\)](#). Other classic findings in psychology suggestive of overreaction include illusion of choice ([Langer, 1975](#)) and illusory correlation (for a review, see [Chapman and Chapman, 1982](#)); [Andreassen \(1987, p. 490\)](#) provides additional references highlighting the tension between under- and overreaction in sequential settings. [Griffin and Tversky \(1992\)](#) argue that the conflicting results can be reconciled if people focus too much on how diagnostic a piece of information is about a given hypothesis but place too little emphasis on the credence of that information. See [Nisbett and Ross \(1980\)](#), especially Chapters 5, 7, and 8) for further discussion.

deterministic component in x_t has been removed, and the variable is demeaned.

2.1.1 Univariate Process

Suppose that x_t follows a linear stationary process

$$x_t = \sum_{\ell=0}^{+\infty} \alpha_\ell \varepsilon_{t-\ell} \quad (1)$$

for some coefficients α_ℓ with $\alpha_0 = 1$ and $\sum_{\ell=0}^{+\infty} \alpha_\ell^2 < +\infty$, and a white noise series of shocks ε_t . While the shocks need not be independent, we assume that they form a martingale difference sequence with respect to the filtration generated by x_t , i.e., $\mathbb{E}_t[\varepsilon_{t+1}] = 0$ where the expectation is conditional on $\{x_t, x_{t-1}, \dots\}$. The class of processes nested by Eq. (1) includes all stationary ARMA processes, with shocks that may exhibit conditional heteroskedasticity. As shown in Section 2.1.2, our approach can be generalized naturally to handle non-stationary x_t provided that the resulting forecast errors are stationary.

We observe an agent making one-step ahead forecasts which are denoted by $\mathbb{F}_t[x_{t+1}]$. We assume that the forecasts are generated as

$$\mathbb{F}_t[x_{t+1}] = b_0 + \sum_{\ell=0}^{+\infty} a_{\ell+1} \varepsilon_{t-\ell}. \quad (2)$$

Here, b_0 is a time-invariant bias term, while the coefficients a_ℓ capture how subjective expectations react to past shocks. If $a_\ell \neq \alpha_\ell$, the subjective reaction to past shocks is different from the reaction of the true process. As shown in Section 3, different models of expectations imply different a_ℓ coefficients. Subjective expectations and observed forecasts need not always coincide (e.g., if the agent is strategic, see Section 3.4), in which case our method reveals biases of the observed forecasts.

We say that the expectation formation process is *unbiased* if $\mathbb{F}_t[x_{t+1}] = \mathbb{E}_t[x_{t+1}]$ with probability one, i.e., the forecast coincides with the true conditional expectation. Furthermore, we say that the agent *underreacts* to shocks that arrived ℓ periods ago (i.e., $\varepsilon_{t-\ell}$) if the perceived response $a_{\ell+1}$ is smaller than the true response $\alpha_{\ell+1}$ in absolute value, i.e., $|a_{\ell+1}| < |\alpha_{\ell+1}|$. The agent is said to *overreact* if $|a_{\ell+1}| > |\alpha_{\ell+1}|$. For example, the agent overreacts to current news, ε_t , if $|a_1| > |\alpha_1|$.

We can express the difference between the true conditional expectation and the ob-

served forecast as

$$\mathbb{E}_{t-1}[x_t] - \mathbb{F}_{t-1}[x_t] = -b_0 - \sum_{\ell=1}^{+\infty} \text{sgn}(\alpha_\ell) b_\ell \varepsilon_{t-\ell}, \quad (3)$$

where $b_\ell \equiv \text{sgn}(\alpha_\ell)(a_\ell - \alpha_\ell)$, $\ell \geq 1$ denotes *bias coefficients*.⁶ This difference can be naturally interpreted as an *ex-ante* forecast error, i.e., forecast error before x_t is realized. For the expectation formation process to be unbiased, all bias coefficients must equal zero. If $\alpha_\ell \neq 0$, the agent overreacts to shocks that arrived ℓ periods ago if $b_{\ell+1} > 0$ and underreacts if $b_{\ell+1} < 0$. For the special case of $\alpha_\ell = 0$, any non-zero bias coefficient indicates overreaction.

Outside experimental settings, we are unlikely to know how x_t is generated. As a result, we typically do not observe either the true conditional expectation or the shocks. The main insight of this paper is that the bias coefficients can be inferred from the behavior of observed forecast errors. To that end, let $e_t \equiv x_t - \mathbb{F}_{t-1}[x_t]$ denote the *forecast error*. Since $x_t = \mathbb{E}_{t-1}[x_t] + \varepsilon_t$ (by virtue of Eq. (1) and $\mathbb{E}_{t-1}[\varepsilon_t] = 0$), Eq. (3) implies

$$\underbrace{x_t - \mathbb{F}_{t-1}[x_t]}_{\equiv e_t} = -b_0 - \sum_{\ell=1}^{+\infty} \text{sgn}(\alpha_\ell) b_\ell \varepsilon_{t-\ell} + \varepsilon_t. \quad (4)$$

As a result, estimating the bias coefficients is equivalent to estimating the *impulse response function* (IRF) of forecast errors. The IRF encapsulates all information about biases in expectation formation. To say whether agents under- or overreact to new information, it is necessary to have some knowledge of the true process—namely, the sign of α_ℓ . For simply testing whether expectations are unbiased, the sign of α_ℓ is not needed. As discussed in Section 2.2, the IRF is inferred from the autocorrelation structure of the observed forecast errors.

Figure 1 illustrates the main ideas behind the measurement framework. The dashed blue line shows the IRF of the true process for x_t . The solid red line plots an example IRF of how the process may be perceived by the agent. As seen in the picture, the bias coefficients b_ℓ are equal to the difference between the two IRFs. Since forecast errors are just the difference between realized values and forecasts, the bias coefficients are in turn equal to the IRF of forecast errors. Our specification of expectations is flexible enough to allow for both under- and overreaction at different lags. In the case of positively autocorrelated x_t (shown in the left panel) we have overreaction for $\ell \in \{1, 2\}$ and underreaction for $\ell \geq 3$.

The right panel of Figure 1 makes it clear why we multiply the bias coefficients

⁶ The sign function, $\text{sgn}(\alpha_\ell)$, is equal to -1 if $\alpha_\ell < 0$ and 1 otherwise.

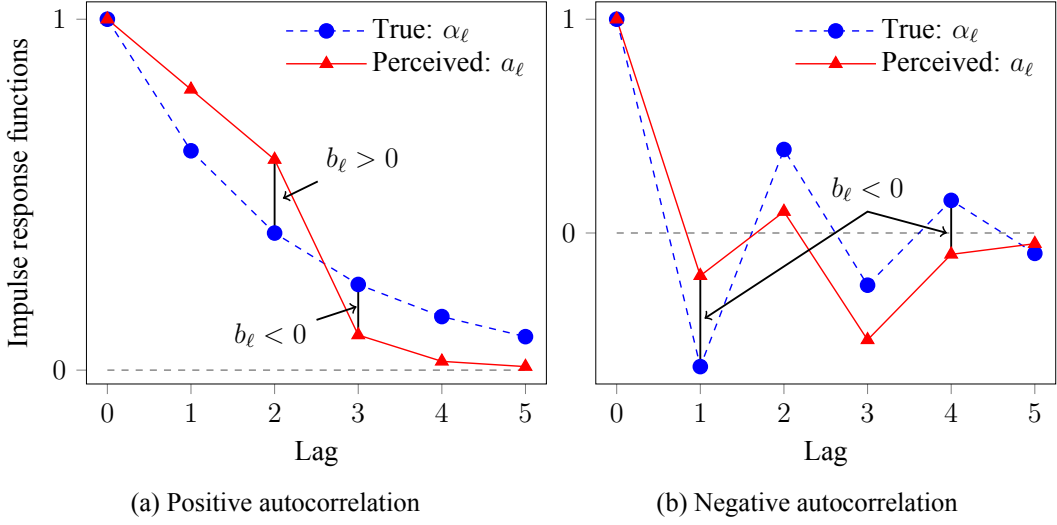


Figure 1: Measurement framework illustrated. The dashed blue line shows the true impulse response function (IRF) of an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$ with $\rho \in (-1, 1)$. The solid red line plots the IRF of the process as it may be perceived by the agent (an example). *Left panel*: Positively autocorrelated process ($\rho > 0$). *Right panel*: Negatively autocorrelated process ($\rho < 0$).

by $\text{sgn}(\alpha_\ell)$. We interpret overreaction to mean that the perceived impulse response is larger than the true impulse response in absolute value. Multiplying by $\text{sgn}(\alpha_\ell)$ ensures that a positive bias coefficient indicates overreaction when the true impulse response is negative. For $\ell = 3$, for example, the perceived impulse response is smaller than the true impulse response, but larger in absolute value, and we classify this bias as overreaction.

2.1.2 Multivariate Process

We now generalize the method to the multivariate case. Consider a mean-zero stochastic process $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{Pt})^\top$. We assume that \mathbf{x}_t follows a linear stationary process

$$\mathbf{x}_t = \sum_{\ell=0}^{+\infty} \Lambda_\ell \boldsymbol{\varepsilon}_{t-\ell}, \quad (5)$$

where $\boldsymbol{\varepsilon}_t \equiv (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Pt})^\top$ is white noise with $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top] = \boldsymbol{\Sigma}$, Λ_ℓ 's are $(P \times P)$ matrices with a typical element given by $\alpha_{ij}^{(\ell)}$, and Λ_0 is the identity matrix. Without loss of generality, we assume that $\boldsymbol{\Sigma}$ is diagonal, so that the shocks in $\boldsymbol{\varepsilon}_t$ are uncorrelated. As before, the shocks need not be independent over time, but we assume $\mathbb{E}_t[\boldsymbol{\varepsilon}_{t+1}] = \mathbf{0}$, so that the shocks form a martingale difference sequence.

Writing out Eq. (5) for a particular variable x_{it} explicitly, we have

$$x_{it} = \sum_{j=1}^P \sum_{\ell=0}^{+\infty} \alpha_{ij}^{(\ell)} \varepsilon_{j,t-\ell}.$$

We generalize the previous specification of expectations in Eq. (2) by assuming that the forecasts are generated as

$$\mathbb{F}_t[x_{i,t+1}] = b_{i0} + \sum_{j=1}^P \sum_{\ell=0}^{+\infty} a_{ij}^{(\ell+1)} \varepsilon_{j,t-\ell}.$$

As before, b_{i0} is a time-invariant bias term, while $a_{ij}^{(\ell+1)}$'s capture the agent's subjective reaction to past shocks. Formally, we say that an agent *overreacts to shock* $\varepsilon_{j,t-\ell}$ if $|a_{ij}^{(\ell+1)}| > |\alpha_{ij}^{(\ell+1)}|$ and *underreacts* if $|a_{ij}^{(\ell+1)}| < |\alpha_{ij}^{(\ell+1)}|$.

Performing the same calculations as in Section 2.1.1, we find that

$$\underbrace{x_{it} - \mathbb{F}_{t-1}[x_{it}]}_{\equiv e_{it}} = -b_{i0} - \sum_{j=1}^P \sum_{\ell=1}^{+\infty} \text{sgn}[\alpha_{ij}^{(\ell)}] b_{ij}^{(\ell)} \varepsilon_{j,t-\ell} + \varepsilon_{it} \quad (6)$$

where $b_{ij}^{(\ell)} = \text{sgn}[\alpha_{ij}^{(\ell)}][a_{ij}^{(\ell)} - \alpha_{ij}^{(\ell)}]$ are the *individual bias coefficients*. As in the univariate case, the expectation formation process is unbiased if and only if all of the individual bias coefficients are zero.

With multiple shocks present, the agent can underreact to some shocks while simultaneously overreacting to others. To obtain a summary measure of bias, it is therefore necessary to combine individual biases. Our approach is to *directly* consider the IRF of forecast errors. Provided that the forecast errors are stationary, by the Wold Representation Theorem (e.g., [Hamilton, 1994](#), Proposition 4.1) we can write

$$x_{it} - \mathbb{F}_{t-1}[x_{it}] = -b_{i0} - \sum_{\ell=0}^{+\infty} \theta_{\ell} \nu_{t-\ell}, \quad (7)$$

for some square-summable coefficients θ_{ℓ} and a white noise series ν_t . The Wold shocks, ν_t , are innovations from a projection of e_{it} on all its past values and hence distinct from the structural shocks, ε_t . Intuitively, we may think of the Wold shocks as a summary measure of “news.” Similarly, we can represent x_{it} as

$$x_{it} = \sum_{\ell=0}^{+\infty} \alpha_{\ell} \xi_{t-\ell},$$

abusing notation somewhat (i.e., α_ℓ are not elements of Λ_ℓ).

With this notation at hand, we define *aggregate bias coefficients* as $b_\ell = \text{sgn}(\alpha_\ell)\theta_\ell$ and say that (for $\alpha_\ell \neq 0$) the agent *overreacts in the aggregate* if $b_\ell > 0$ and *underreacts in the aggregate* if $b_\ell < 0$. (If $\alpha_\ell = 0$ and $\theta_\ell \neq 0$, the agent is said to overreact.) Note that for Eq. (7) to be valid, it is only necessary that the forecast errors are stationary, a significantly weaker requirement than x_t being stationary. Hence, the approach outlined in Section 2.1.1 generalizes naturally to the multivariate case, and it can handle forecasts of non-stationary variables provided that the forecast errors are stationary.

For the expectations formation process to be unbiased, it is necessary (but *not* sufficient) for the aggregate bias coefficients to be zero. To see this, suppose that some aggregate bias coefficients are not zero but—to obtain a contradiction—all individual bias coefficients are zero. In that case, Eq. (6) implies that $x_{it} - \mathbb{F}_{t-1}[x_{it}] = \varepsilon_{it}$. But then all aggregate bias coefficients are zero, a contradiction. To see that zero aggregate bias coefficients are not sufficient for the expectation formation process to be unbiased, suppose that $x_{it} = 0$ with probability one but $\mathbb{F}_t[x_{it}] = \xi_t$ where ξ_t is an i.i.d. shock. Then, the agent overreacts to ξ_t at the individual level, but the aggregate bias coefficients are all zero since ξ_t is an i.i.d. shock.

In general, the relationship between the individual and aggregate bias coefficients is non-linear.⁷ Let γ_k denote the autocovariance function of forecast errors; the autocovariance function can be calculated from Eq. (6). Then, the θ_ℓ coefficients satisfy

$$\sum_{\ell=0}^{+\infty} \theta_\ell \theta_{\ell+k} = \frac{\gamma_k}{\gamma_0} \sum_{\ell=0}^{+\infty} \theta_\ell^2 \text{ for } k = 1, 2, \dots \quad (8)$$

Suppose that the autocovariances vanish at some finite lag K . Then, the process for the forecast errors has a finite Wold representation, and Eq. (8) boils down to a set of K nonlinear equations that can be solved to find θ_ℓ 's (see, e.g., [Ansley, Spivey, and Wroblecki, 1977](#)).

Figure 2 illustrates the relationship between the individual and aggregate bias coefficients. In the example given, there are two shocks that have equal variances. The agent overreacts to the first shock but underreacts to the second shock. The figure shows the two sets of individual bias coefficients as well as the implied aggregate bias coeffi-

⁷ It is well known that aggregation of simple stochastic processes can lead to complicated behavior. As a result, it is very difficult to obtain a characterization that goes beyond Eq. (8). For example, adding up P independent AR(1) processes leads to an ARMA($P, P-1$) process for the sum ([Granger, 1980](#)). The literature on aggregation typically studies the limiting behavior as $P \rightarrow +\infty$. In the present case, we think of P as being small to moderate. While the impulse responses do not aggregate linearly, it is well known that autocorrelations of mutually-uncorrelated processes do aggregate in a straightforward manner (see, e.g., [Anderson, 1975](#)).

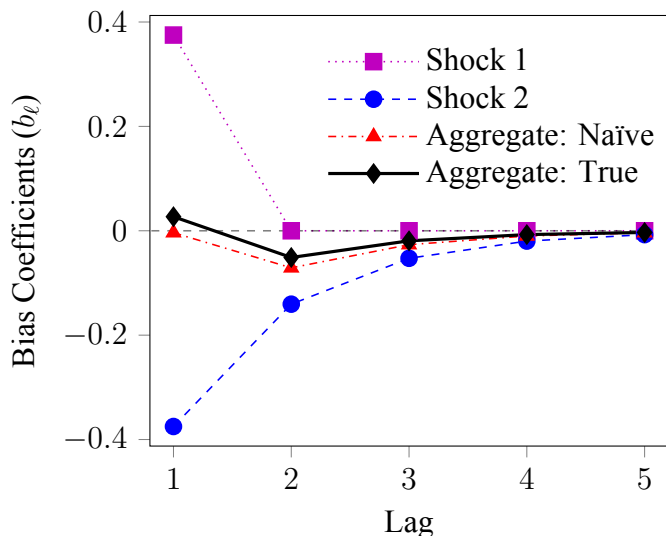


Figure 2: An example of individual and aggregate bias coefficients. There are two shocks with equal variances. The individual bias coefficients are shown by the dotted magenta and dashed blue lines. The aggregate bias coefficients are given by the solid black line. The dash-dotted red line plots a naïve estimate of the aggregate bias coefficients, given by the weighted average of the two individual bias coefficients; the weights are fractions of total variance explained by each shock (γ_{j0}/γ_0). The aggregate bias coefficients are found by solving Eq. (8) numerically and taking the invertible solution. The values for the individual bias coefficients are those implied by (i) diagnostic expectations of [Bordalo, Gennaioli, and Shleifer \(2018b\)](#); and (ii) sticky information model of [Mankiw and Reis \(2002\)](#) in a simple univariate setting; see Figure 3 for the parameter values.

coefficients, obtained by solving Eq. (8). Not surprisingly, the aggregate bias coefficients lie between the two individual bias coefficients. The first aggregate coefficient is positive, implying overreaction in the aggregate, while the remaining aggregate bias coefficients are negative.

The figure also plots a *naïve estimate* of the aggregate bias coefficients. The naïve estimate is given by a simple weighted average of the two individual bias coefficients; the weights are the fractions of the variance in forecast errors explained by each shock. We see that—in this example—the aggregate bias coefficients are well approximated by a weighted average of the individual bias coefficients. The example also illustrates an important limitation of our method. If the biases of the agent “cancel out,” our method may fail to detect bias in the aggregate. Again, aggregate bias coefficients being zero is a necessary, but not sufficient, condition for the expectation formation process to be unbiased.

We conclude this section by noting that in many empirical settings, it is difficult to go beyond the estimation of aggregate bias coefficients. One basic challenge is that we may not observe the forecasts of all the variables in \mathbf{x}_t . However, even if we did observe all of the forecasts, the individual bias coefficients cannot be recovered by simply estimating

the (multivariate) IRF of forecast errors. The reason is due to the standard identification problem in multivariate models. If we estimate the IRF of the forecast errors, the estimated coefficients only measure the response of forecast errors to reduced-form innovations. Economically, this IRF may not be of direct interest. For example, we may be interested to learn how inflation expectations respond to a monetary policy shock. To measure this response, as is well known, additional assumptions are necessary (for a recent discussion, see [Ramey, 2016](#)).

2.2 Estimation

If we could observe the true shocks, it would be straightforward to estimate the bias coefficients by simply regressing forecast errors $e_t = x_t - \mathbb{F}_{t-1}[x_t]$ on past shocks. For instance, in the case of a single shock, one would estimate

$$e_t = \alpha + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_K \varepsilon_{t-K} + u_t. \quad (9)$$

The estimated bias coefficients for $\ell \geq 1$ are then $\hat{b}_\ell = -\text{sgn}(\alpha_\ell) \hat{\beta}_\ell$. As we discuss in [Section 2.3](#), this approach is followed in several prominent existing studies, sometimes implicitly.

In many cases of interest, however, we do not have enough information to estimate the true shocks. In these situations we can still recover the bias coefficients by estimating the IRF of the forecast errors. To estimate the IRF flexibly, we use the method of *local projections* ([Jordá, 2005](#)). For each $s = 0, 1, \dots, (L - 1)$, we estimate the following regression by least squares:

$$e_{t+s} = \alpha^{(s)} + \beta_1^{(s)} e_{t-1} + \beta_2^{(s)} e_{t-2} + \cdots + \beta_K^{(s)} e_{t-K} + u_t. \quad (10)$$

Here, K denotes the number of lagged forecast errors included in the local projection. The estimated bias coefficients for $\ell \geq 1$ are then $\hat{b}_\ell = -\text{sgn}(\alpha_\ell) \hat{\beta}_1^{(\ell-1)}$. The time-invariant bias coefficient b_0 is estimated by simply calculating the sample average of e_t .

While our preferred estimation method is local projections, it is possible to estimate the bias coefficients by fitting a high-order moving average model using maximum likelihood. In our experience, when maximum likelihood estimation works numerically, the resulting point estimates are very similar; an example is provided in [Section 6](#). However, there are major benefits to using local projections. First, it is immediate to extend our method to cases in which we have multiple forecasters or want to pool multiple forecasts (i.e., when we have panel data). Second, it is straightforward to adjust standard

errors for clustering that occurs when individual forecasts are used (Keane and Runkle, 1990). Finally, maximum likelihood estimation of high-order moving average models can be numerically challenging. An alternative that may be useful when statistical power is otherwise limited is to estimate the IRF by fitting a more tightly parametrized time series model (e.g., an AR(4) for quarterly forecast errors).

Intuition. Some readers may still wonder how we can recover the bias coefficients if the ε_t 's in Eq. (4) are unobserved. The bias coefficients are identified from the structure of autocorrelation in the forecast errors. Suppose that the true process for x_t is positively autocorrelated but the agent underreacts to new information. In that case, the forecast errors will be positively autocorrelated. With overreaction, in contrast, we would see negatively autocorrelated forecast errors. Implicitly, the method in Eq. (10) maps the observed autocorrelations into an implied IRF.

To see this in more detail, suppose that x_t follows an AR(1) process $x_t = \rho x_{t-1} + \varepsilon_t$, $\rho \in (-1, 1)$, and consider the following two models of expectations, one exhibiting underreaction and one exhibiting overreaction. (The precise descriptions of these models are given in Section 3.) For the sticky information model proposed by Mankiw and Reis (2002), forecast errors turn out to also follow an AR(1) process (see Section 3.2):

$$e_t = \lambda \rho e_{t-1} + \varepsilon_t.$$

Here $\lambda \in [0, 1]$ is a parameter measuring the stickiness of expectations. If the process is positively autocorrelated ($\rho > 0$), forecast errors are also positively autocorrelated. The IRF of forecast errors decays geometrically.

Now, in contrast, suppose that expectations are diagnostic (Bordalo, Gennaioli, and Shleifer, 2018b). In this case, forecast errors follow an MA(1) process (see Section 3.3):

$$e_t = \varepsilon_t - \theta \rho \varepsilon_{t-1},$$

where $\theta \geq 0$ is a parameter capturing the extent to which agents overweight representative events. Therefore, forecast errors are negatively correlated at the first lag. At all other lags, forecast errors are uncorrelated. The IRF of forecast errors is equal to $(-\theta\rho)$ at the first lag and zero at all subsequent lags.

2.3 Related Work

We now discuss how our framework relates to existing work.

In the paper that introduced rational expectations, Muth (1961) already had a section on “Deviations from Rationality” (p. 321). In that section Muth studied a specification

of expectations (his equation 3.18) that is a special case of our Eq. (2). In the notation of the present model, Muth allowed the subjective reaction to current news (a_1) to differ from the true reaction of the process (α_1). In the present paper we continue in Muth's steps by allowing the subjective reaction to news in *all* periods to be different from the true reaction of the process. In addition, we show how to estimate the bias coefficients empirically.

There is a large literature that tests whether forecasts are rational by looking at whether forecast errors are predictable, as surveyed by [Pesaran and Weale \(2006, Section 5\)](#). In fact, our main regression specification, Eq. (10), is a special case of such tests, with the predictors given by past forecast errors. A major strength of these tests is that they are relatively assumption free. However, the traditional tests can reject the null hypothesis without being particularly informative about the alternative.⁸ In contrast, rejections of our test have a natural economic interpretation in terms of under- or overreaction to news at various lags.

Broadly speaking, existing empirical approaches to measuring biases in expectation formation tend to impose assumptions on either (i) the underlying process; or (ii) the way expectations are formed. In the context of our framework, these assumptions correspond to Eqs. (1) and (2) and their multivariate generalizations.

In category (i), this paper is most closely related to [Coibion and Gorodnichenko \(2012\)](#). Similarly to our work, they study how forecast errors respond to shocks. While the authors do not frame their theoretical discussion in terms of bias coefficients, they derive how the forecast errors react to shocks in a number of models of expectations. These calculations are the same as those that we perform in our paper. However, in their empirical work Coibion and Gorodnichenko directly estimate the shocks using methods that are somewhat specific to their setting (expectations of macroeconomic variables). Our method has weaker informational requirements and may be more portable across applications. Nevertheless, when the restrictions made to estimate the shocks are in fact satisfied, the procedure employed by Coibion and Gorodnichenko typically yields more precise estimates (see Section 5).⁹

A prominent literature on expectations of stock returns also falls into category (i). At short horizons, classic asset pricing theory predicts that stock prices should follow

⁸ For example, [Coibion and Gorodnichenko \(2015, p. 2651\)](#) write that “[...] when traditional tests identify a rejection of the null hypothesis of FIRE [full information rational expectations], this rejection is not directly informative about other theories of the expectations formation process in the absence of a clear theoretical mapping from the theory to the empirical tests.”

⁹ An advantage of our method is that it uses all of the variation in x_t . In contrast, it may only be possible to measure a fraction of the variation in the true shocks in practice. If only a small fraction of the variation in the true shocks is used to apply the Coibion-Gorodnichenko procedure, our method may well lead to more precise estimates.

martingales, and therefore changes in stock prices should be unpredictable (see, e.g. [Cochrane, 2005](#), p. 22). Hence, a reasonable first approximation for stock returns is that $x_t = \varepsilon_t$, where ε_t is a martingale difference sequence. In that case, the IRF representation in Eq. (4) implies that

$$\mathbb{F}_{t-1}[x_t] = b_0 + \sum_{\ell=1}^{+\infty} b_\ell x_{t-\ell}.$$

As a result, the bias coefficients can be recovered by simply regressing current expectations on past stock returns. The procedure is likely to yield more precise estimates than our method. Extrapolative expectations of stock returns have been documented by [Graham and Harvey \(2001\)](#), [Vissing-Jorgensen \(2004\)](#), [Dominitz and Manski \(2011\)](#) and [Greenwood and Shleifer \(2014\)](#), among others, and are often interpreted as prima facie evidence for overreaction. Our framework highlights that stock returns must be uncorrelated over the relevant time horizon for the above specification to yield a meaningful measure of overreaction. If stock returns are correlated over time, some degree of extrapolation is justified.¹⁰

There is also a large experimental literature on expectations. In experimental settings, the experimenter generates the data and hence can directly observe the true shocks. The experimental literature dates back to at least [Schmalensee \(1976\)](#); Schmalensee cites [Fisher \(1962\)](#) as a source for his experimental design, but the reference by Fisher is difficult to obtain. A recent experimental study that is particularly relevant for the present paper is [Ma, Landier, and Thesmar \(2017\)](#). Ma, Landier, and Thesmar estimate an empirical model of expectation formation that is remarkably stable across various treatments; they also show how expectations respond to a unit shock in the model they consider (their Figure 5).

An important example of a category (ii) paper is given by [Coibion and Gorodnichenko \(2015\)](#). In that paper, the authors show that regressing forecast errors on lagged forecast revisions recovers structural parameters in a number of models featuring underreaction. Importantly, it is not necessary to know the data-generating process to recover parameters governing expectation formation. However, making the link between empirical estimates and structural parameters requires assumptions about how expectations are formed. If these assumptions are violated, it may not be straightforward to interpret the empirical estimates, analogously to the stock market case discussed above. Once again, if the assumptions are in fact satisfied, the method typically yields more efficient estimates than our procedure.

¹⁰ [Cutler, Poterba, and Summers \(1990a\)](#) document that in a sample of 13 developed countries over 1960–1988, the autocorrelation of yearly excess aggregate stock market returns is roughly 0.02 (s.e. = 0.004).

In this paper, we focus on point forecasts. These, of course, may be forecasts of higher moments of the data. For example, x_t could represent the volatility of stock market returns in a given quarter. In some empirical contexts, however, we have access to more information, including density forecasts. This additional information can be used to develop additional tests, as recently shown by [Augenblick and Lazarus \(2018\)](#) and [Augenblick and Rabin \(2018\)](#). An important advantage of the tests developed by Augenblick, Lazarus, and Rabin is that they do not require data on realizations, as the authors proceed by characterizing the amount of movement in beliefs that is permissible under Bayesian updating.

3 Mapping Existing Models

We now show how existing models of expectations can be mapped into our framework. The exercise leads to two key takeaways. First, our framework is flexible enough to nest all major models of expectations as special cases. Second, the implied bias coefficients provide a useful lens for looking at models of expectations. For instance, some models that may intuitively be thought to only exhibit underreaction, in fact, imply overreaction at some lags.

To obtain closed-form expressions, we assume that x_t follows a stationary AR(1) process:

$$x_t = \rho x_{t-1} + \varepsilon_t, \rho \in (-1, 1),$$

where—as in Section 2— ε_t is a white noise error term with $\mathbb{E}_t[\varepsilon_{t+1}] = 0$. The simple AR(1) process is a reasonable first approximation for many economic time series. For more complicated processes, it is often difficult to obtain the bias coefficients analytically. In those cases, it is still straightforward to obtain estimates by simulation.

To streamline the exposition, we discuss four major models of expectations in the main text (rational expectations, sticky information, diagnostic expectations, and adjustment costs). In Appendix A, we show how to obtain bias coefficients for models of adaptive expectations, noisy information, misperceived law of motion, extrapolative expectations, learning, and overconfidence.

The results for the models discussed in the current section are summarized in Figure 3. As shown in the figure, existing models have sharp predictions for the structure of bias coefficients. The sticky information model by [Mankiw and Reis \(2002\)](#) implies that the bias coefficients are negative and decay geometrically (dashed blue line). In contrast, diagnostic expectations of [Bordalo, Gennaioli, and Shleifer \(2018b\)](#) predict that the agent overreacts to current news but reacts rationally to all past news (dotted

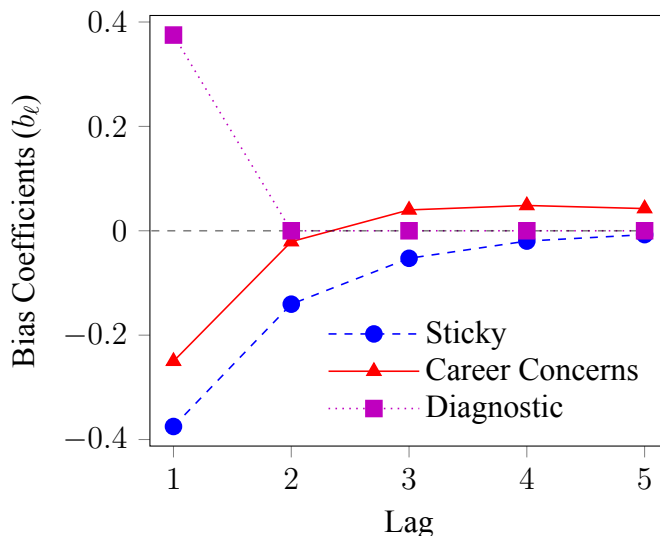


Figure 3: Bias coefficients for selected models of expectations. Positive bias coefficients indicate overreaction to news at a particular lag, and negative coefficients indicate underreaction. Unbiased reaction to news is given by a zero bias coefficient. The underlying process for x_t is $x_t = 0.75x_{t-1} + \varepsilon_t$. The models shown are: (i) sticky information model of [Mankiw and Reis \(2002\)](#) with $\lambda = 0.50$; (ii) forecasting with adjustment costs (Section 3.4) with $\alpha = 1.0$ and $\delta = 0.50$; and (iii) diagnostic expectations of [Bordalo, Gennaioli, and Shleifer \(2018b\)](#) with $\theta = 0.50$.

magenta line). Finally, we plot the bias coefficients for a model in which the agent is rational but must incur a quadratic adjustment cost to change forecasts from one period to the next (solid red line). These adjustment costs may be interpreted as representing career concerns (in an admittedly reduced-form fashion). At the chosen parameter values, the model with adjustment costs predicts strong underreaction to current news but mild overreaction to news received further in the past. The result illustrates how bias coefficients can be helpful in understanding models of expectations.

3.1 Rational Expectations

We now show how to map four specific models into our framework (see Appendix A for additional models). Rational expectations in the sense of [Muth \(1961\)](#) are given by

$$\begin{aligned} a_\ell &= \rho^\ell \\ b_\ell &= 0 \end{aligned}$$

Here, the perceived response to a shock, a_ℓ , is identical to the true response of x_t . As a result, all bias coefficients are zero.

3.2 Sticky Information

Consider the sticky information model of expectations proposed by [Mankiw and Reis \(2002\)](#); see also [Carroll \(2003\)](#). Each period a fraction $(1 - \lambda) \in (0, 1]$ of agents update their forecast to the full-information rational expectation. The remaining agents use information obtained in some previous period to form expectations that are rational conditional on their information set.¹¹ Given these assumptions, expectations at the aggregate (or consensus) level follow

$$\mathbb{F}_t[x_{t+1}] = (1 - \lambda) \sum_{\ell=0}^{+\infty} \lambda^\ell \mathbb{E}_{t-\ell}[x_{t+1}]. \quad (11)$$

For the AR(1) model, we have that $\mathbb{E}_{t-\ell}[x_{t+1}] = \rho^{\ell+1}x_{t-\ell}$, and some algebra yields

$$\mathbb{F}_t[x_{t+1}] = \sum_{\ell=0}^{+\infty} \rho^{\ell+1}(1 - \lambda^{\ell+1})\varepsilon_{t-\ell}. \quad (12)$$

As a result, we find that

$$\begin{aligned} a_\ell &= \rho^\ell(1 - \lambda^\ell) \\ b_\ell &= -\operatorname{sgn}(\rho^\ell)(\lambda\rho)^\ell \end{aligned}$$

As long as expectations do not adjust to news immediately ($\lambda > 0$), the sticky information model exhibits underreaction at all lags.¹²

As seen above, the bias coefficients depend on both (i) how people form expectations; and (ii) the data-generating process. Different processes for x_t will imply different bias coefficients, even if people form expectations in the same way. In the present example, the bias coefficients are larger in absolute value if the process is more persistent. Intuitively, underreaction is more severe when the process is highly persistent.

3.3 Diagnostic Expectations

Suppose that the agent has diagnostic expectations as in [Bordalo, Gennaioli, and Shleifer \(2018b\)](#) and overweights representative events. [Bordalo, Gennaioli, and Shleifer \(2018b\)](#),

¹¹ [Reis \(2006, Section 5\)](#) provides a microfoundation for the Poisson adjustment process.

¹² Underreaction at all lags with sticky information extends to more general processes. To see this, consider a general x_t as in Eq. (1) and perform the same calculations as for the AR(1) case. The calculation shows that in the general case $b_\ell = -\operatorname{sgn}(\alpha_\ell)\lambda^\ell\alpha_\ell$.

Proposition 1) show that in this case expectations follow

$$\mathbb{F}_t[x_{t+1}] = \mathbb{E}_t[x_{t+1}] + \theta \{\mathbb{E}_t[x_{t+1}] - \mathbb{E}_{t-1}[x_{t+1}]\}, \theta \geq 0, \quad (13)$$

where θ is a parameter capturing the extent to which the agent overweights representative events. The expression can be rewritten as $\mathbb{F}_t[x_{t+1}] = \mathbb{E}_t[x_{t+1}] + \rho\theta\varepsilon_t$. Therefore, diagnostic expectations imply that

$$a_\ell = \begin{cases} \rho(1 + \theta) & \text{if } \ell = 1 \\ \rho^\ell & \text{if } \ell \geq 2 \end{cases} \quad \text{and} \quad b_\ell = \begin{cases} \theta|\rho| & \text{if } \ell = 1 \\ 0 & \text{if } \ell \geq 2 \end{cases}$$

Hence, diagnostic expectations predict overreaction to current news and unbiased reaction to all other news.¹³

3.4 Adjustment Costs

Finally, we assume the agent has rational expectations but faces a cost in adjusting forecasts from one period to the next. We interpret the adjustment cost as a stand-in for reputational costs or career concerns. For example, forecasters who change their forecasts by large amounts may be perceived as having lower forecasting ability.

Similarly to [Coibion and Gorodnichenko \(2015, p. 2660\)](#), suppose that in each period t the agent makes a forecast of x_{t+1} . The agent wishes to minimize the mean-squared error of the prediction but faces a quadratic adjustment cost.¹⁴ Denoting the current value of x_t as x and the previous forecast by F , the Bellman equation of the agent is given by

$$V(x, F) = \min_{F'} \frac{1}{2} \mathbb{E}[(\rho x + \tilde{\varepsilon} - F')^2] + \frac{\alpha}{2} (F' - F)^2 + \delta \mathbb{E}[V(\rho x + \tilde{\varepsilon}, F')],$$

where F' is the current period's forecast, $\alpha \geq 0$ is the weight on the adjustment cost, $\delta \in (0, 1)$ is a discount factor, and we have used tildes to denote random variables. The

¹³ The prediction is conditional on the underlying process being an AR(1). A distinctive feature of diagnostic expectations is that they depend on the structure of the underlying process, i.e., Eq. (13) is endogenous to the model. In the sticky information model, parameter λ is also likely to be endogenous to the process.

¹⁴ The setup is different from a situation in which an agent does not want to deviate from the consensus forecast, leading to game-theoretic considerations. This alternative situation is fully characterized by [Coibion and Gorodnichenko \(2012, pp. 126–129\)](#) who show—similarly to the model with sticky information—that the forecast errors follow an AR(1) process, implying geometrically decaying bias coefficients.

first-order condition is

$$-(\rho x - F') + \alpha(F' - F) + \delta \mathbb{E}[V_F(\rho x + \tilde{\varepsilon}, F')] = 0.$$

The envelope condition is just $V_F(x, F) = -\alpha(F' - F)$. Therefore, the optimal forecast is given by

$$(F')^* = \frac{\rho x + \alpha(1 - \delta)F}{1 + \alpha(1 - \delta)}.$$

If the agent is fully patient ($\delta = 1$) or there is no adjustment cost ($\alpha = 0$), the forecast coincides with the true conditional expectation. In the other extreme, if $\alpha \rightarrow +\infty$, then it is optimal to never change the forecast.

Defining $\phi \equiv \alpha(1 - \delta)/[1 + \alpha(1 - \delta)]$, the optimal forecasting rule is

$$\mathbb{F}_t[x_{t+1}] = (1 - \phi)\rho x_t + \phi \mathbb{F}_{t-1}[x_t].$$

Performing similar manipulations to those in Appendix A.1, we arrive at

$$\mathbb{F}_t[x_{t+1}] = (1 - \phi)\rho \sum_{\ell=0}^{+\infty} \left[\frac{\phi^{\ell+1} - \rho^{\ell+1}}{\phi - \rho} \right] \varepsilon_{t-\ell}.$$

Hence, the bias coefficients are equal to

$$\begin{aligned} a_\ell &= (1 - \phi)\rho \left[\frac{\phi^\ell - \rho^\ell}{\phi - \rho} \right] \\ b_\ell &= \text{sgn}(\rho^\ell)(a_\ell - \rho^\ell) \end{aligned}$$

A first intuition may be that adjustment costs can only generate underreaction. However, inspecting the expressions above, it is clear that this is not the case. Indeed, the bias coefficients for this model in Figure 3 are positive for $\ell \geq 2$ at some parameter values. Hence, the model with adjustment costs may generate overreaction (to more distant news).

The example also shows that the first-order autocorrelation of forecast errors can be misleading as a measure of underreaction. For the present model, the first-order autocorrelation is equal to

$$\text{Corr}(e_t, e_{t-1}) = \frac{\phi\rho[1 + \phi + \rho(1 - \phi)]}{1 + \rho + \phi(1 - \rho)[1 + \rho(1 - \phi)]}.$$

Since $\phi \in [0, 1)$, if x_t is positively autocorrelated, the first-order autocorrelation of forecast errors is always positive. Nevertheless, as already discussed, expectations may

exhibit overreaction. Hence, the first-order autocorrelation can be misleading as a measure of underreaction. For example, at the parameter values used in Figure 3, the autocorrelation is equal to roughly 0.24, even though expectations overreact to news in multiple periods. Ignoring potential overreaction in later periods, as is commonly done in existing empirical work, can lead to incorrect inferences.

4 Extensions

We now discuss three extensions to the basic framework: multiple-step-ahead forecasts, heterogeneity, and measurement error.

4.1 Multiple-Step-Ahead Forecasts

Forecasters commonly make multiple-step-ahead forecasts. For instance, we may observe an analyst making one-year-ahead inflation forecasts every quarter. It is straightforward to modify our methodology to account for such cases.

Suppose that the agent makes h -step ahead forecasts with $h \geq 1$ denoting the forecast horizon. To simplify notation, suppose that x_t follows a univariate process as given by Eq. (1). We assume that the forecasts are generated as

$$\mathbb{F}_t[x_{t+h}] = b_0 + \sum_{\ell=0}^{+\infty} a_{\ell+h} \varepsilon_{t-\ell}.$$

Performing the same calculations as in Section 2.1 shows that

$$x_t - \mathbb{F}_{t-h}[x_t] = -b_0 - \sum_{\ell=h}^{+\infty} \text{sgn}(\alpha_\ell) b_\ell \varepsilon_{t-\ell} + \sum_{\ell=0}^{h-1} \alpha_\ell \varepsilon_{t-\ell}.$$

Comparing the equation above to Eq. (4), we observe an additional term stemming from the multiple-step ahead nature of forecasts. Even if subjective expectations react to news in an unbiased way, forecast errors are mechanically autocorrelated—up to lag $(h-1)$ —if the underlying process is autocorrelated at these lags. The result is well known (see, e.g., [Diebold and Lopez, 1996](#)).

The methodology again boils down to estimating the IRF of forecast errors. Differently from the one-step-ahead case, the first $(h-1)$ impulse responses need to be discarded. The remaining impulse responses are converted to bias coefficients, with $b_{\ell+h}$ giving the biased reaction to news that arrived ℓ periods ago.

4.2 Heterogeneity and Aggregation

Existing research has documented that expectations are heterogeneous across individuals (e.g., [Manski, 2004](#), Section 5). Our method can easily be applied to subsamples of the population. For example, we may estimate the bias coefficients for young and old forecasters. When the method is applied to the whole population, it recovers an average of the bias coefficients of the individual forecasts, as we now show.

Suppose that, as in Section 2.1.2, x_t follows a multivariate process

$$x_t = \sum_{j=1}^P \sum_{\ell=0}^{+\infty} \alpha_j^{(\ell)} \varepsilon_{j,t-\ell},$$

where we have dropped one subscript to simplify the notation. The forecast of agent i is generated as

$$\mathbb{F}_{it}[x_{t+1}] = b_{i0} + \sum_{j=1}^P \sum_{\ell=0}^{+\infty} a_{ij}^{(\ell+1)} \varepsilon_{j,t-\ell}, \quad i = 1, 2, \dots, N,$$

with i indexing individual forecasters (not different variables, as in Section 2.1.2). Denote the consensus (average) forecast by $\mathbb{F}_t[x_{t+1}] \equiv \frac{1}{N} \sum_i \mathbb{F}_{it}[x_{t+1}]$. Then, we see that the consensus forecast is given by

$$\mathbb{F}_t[x_{t+1}] = \bar{b}_0 + \sum_{j=1}^P \sum_{\ell=0}^{+\infty} \bar{a}_j^{(\ell+1)} \varepsilon_{j,t-\ell},$$

where $\bar{b}_0 \equiv \frac{1}{N} \sum_i b_{i0}$ and $\bar{a}_j^{(\ell)} \equiv \frac{1}{N} \sum_i a_{ij}^{(\ell)}$. Applying the results from Section 2.1.2, the individual bias coefficients of the consensus forecast, $\bar{b}_j^{(\ell)}$, are equal to

$$\bar{b}_j^{(\ell)} = \text{sgn}[\alpha_j^{(\ell)}][\bar{a}_j^{(\ell)} - \alpha_j^{(\ell)}] = \text{sgn}[\alpha_j^{(\ell)}] \left[\frac{1}{N} \sum_{i=1}^N \{a_{ij}^{(\ell)} - \alpha_j^{(\ell)}\} \right] = \frac{1}{N} \sum_{i=1}^N b_{ij}^{(\ell)},$$

where $b_{ij}^{(\ell)} \equiv \text{sgn}[\alpha_j^{(\ell)}][a_{ij}^{(\ell)} - \alpha_j^{(\ell)}]$. Therefore, with heterogeneity, the bias coefficients of the consensus forecast equal the average bias coefficients of the individual forecasts. As a result, the estimated bias coefficients are likely to be similar irrespective of whether consensus or individual forecasts are used.

We emphasize that the linear aggregation result only holds for the *individual* bias coefficients, in the sense defined in Section 2.1.2. For the aggregate bias coefficients, aggregation is non-linear, and therefore the aggregate bias coefficients of the consensus forecast are not given by a linear combination of the aggregate bias coefficients of the

individual forecasts.

4.3 Measurement Error

In practice, we may only be able to measure subjective expectations with some error. Such error could arise because of imperfect measurement (e.g., survey questions may be confusing). Measurement error could also be given an economic interpretation if subjective expectations are inherently random. A concrete example of the latter is provided in Appendix A (Section A.6) that discusses a stylized model of overconfidence.

Measurement error tends to mask any existing predictability of the forecast errors, biasing the estimated bias coefficients towards zero. As a result, the empirically estimated bias coefficients are likely to represent a lower bound on the true bias coefficients. Measurement error is less of an issue in datasets in which many individual forecasts are available. For these datasets, averaging across forecasters reduces measurement error by virtue of the Law of Large Numbers (assuming that measurement errors across forecasters are not too dependent).¹⁵

In Appendix B.1 we provide explicit formulas for the attenuation bias caused by measurement error for two models of expectations (sticky information and diagnostic expectations) when x_t follows an AR(1). The attenuation bias can be substantial, underscoring the benefits of datasets with a large cross-sectional dimension.

While our method is sensitive to measurement error, so are most other existing techniques. For instance, a common test for the optimality of expectations with respect to some information set estimates

$$x_t = \alpha + \beta \mathbb{F}_{t-1}[x_t] + u_t$$

and then tests the joint restriction $\alpha = 0$ and $\beta = 1$. The test goes back to at least [Mincer and Zarnowitz \(1969\)](#), p. 9, who also credit Henri Theil, see p. 5); for a more detailed discussion, see [Diebold and Lopez \(1996\)](#). Clearly, even if $\alpha = 0$ and $\beta = 1$ in the equation above, if we only observe a noisy version of $\mathbb{F}_{t-1}[x_t]$, the estimate of β will be biased towards zero. For the method of [Coibion and Gorodnichenko \(2015\)](#), measurement error can be even more pernicious and may even change the estimated sign, as shown in Appendix B.2.

¹⁵ Existing literature has developed techniques for consistently estimating time series models in the presence of measurement error, see [Staudenmayer and Buonaccorsi \(2005\)](#) and references therein. It may be possible to combine these techniques with the approach taken in our paper to obtain consistent estimates of the bias coefficients in the presence of measurement error. However, pursuing this path is outside the scope of the present paper.

5 Monte Carlo Experiment

To further investigate the validity of our methodology, we perform a Monte Carlo experiment. The purpose of the experiment is twofold. First, we gauge the statistical power of our procedure. Second, we investigate the performance of our method when some of the underlying assumptions are violated.

To facilitate comparison, the baseline design follows [Coibion and Gorodnichenko \(2012, Appendix B\)](#). The true process is an AR(1) calibrated to match the key features of quarterly GDP deflator inflation:

$$x_t = \rho x_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2), t = 1, 2, \dots, T.$$

We set $\rho = 0.85$, $\sigma_\varepsilon^2 = 1.005$ and use a sample size of $T = 150$ periods. The value for x_1 is drawn from the stationary distribution of x_t , i.e., $\mathcal{N}(0, \sigma_\varepsilon^2/(1 - \rho^2))$.

Subjective expectations follow the sticky information model of [Mankiw and Reis \(2002\)](#):

$$\mathbb{F}_t[x_{t+1}] = (1 - \lambda)\rho x_t + \lambda\rho \mathbb{F}_{t-1}[x_t].^{16}$$

The level of information stickiness is $\lambda = 0.75$. We set $\mathbb{F}_0[x_1] = 0$ to start up the recursion. To allow for measurement error, the *observed* forecasts, $\mathbb{F}_t^*[x_{t+1}]$, are equal to the true forecasts plus some noise v_t (independent of the true shocks):

$$\mathbb{F}_t^*[x_{t+1}] = \mathbb{F}_t[x_{t+1}] + v_t, v_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_v^2). \quad (14)$$

As discussed in [Section 4.3](#), measurement error could have multiple economic interpretations, including data collection problems (e.g., poorly constructed surveys) as well as expectations that are inherently random (e.g., overconfidence, see [Section A.6](#)).

We parametrize the size of measurement error by the R^2 of [Eq. \(14\)](#). From [Eq. \(12\)](#), the fraction of the variation in the observed forecasts that is explained by the true forecasts is

$$R^2 = \frac{\kappa\sigma_\varepsilon^2}{\kappa\sigma_\varepsilon^2 + \sigma_v^2} \text{ where } \kappa = \frac{(1 - \lambda)^2\rho^2(1 + \lambda\rho^2)}{(1 - \rho^2)(1 - \lambda\rho^2)(1 - \lambda^2\rho^2)}. \quad (15)$$

In [Appendix B.2](#), we show that the existing evidence on inflation expectations points to $\sigma_v \approx 0.21$, implying an R^2 of 0.95 at the chosen parameter values. Since other empirical settings may have more measurement error, we provide the results for a range of values for R^2 . To avoid potential initial value effects, we simulate the model for 1,150 periods and discard the first 1,000 periods. The remaining 150 periods are then

¹⁶ To obtain this recursion, use [Eq. \(21\)](#), to find that $\mathbb{F}_t[x_{t+1}] - \lambda\mathbb{F}_{t-1}[x_{t+1}] = (1 - \lambda)\rho x_t$ and then substitute $\mathbb{F}_{t-1}[x_{t+1}] = \rho\mathbb{F}_{t-1}[x_t]$ as derived in [Appendix B.1 \(Section B.2\)](#).

used for estimation.

The results are shown in Figure 4. The left column plots the estimated bias coefficients when the true shocks are observed. This estimation procedure regresses the forecast errors on the true shocks, as given in Eq. (9), and is essentially identical to the method proposed by Coibion and Gorodnichenko (2012).¹⁷ The top row has $R^2 = 1$ (no measurement error), and the lower rows include progressively more measurement error. The shaded areas give 90% of the Monte Carlo realizations.

The estimates obtained using the true shocks are unbiased. Measurement error does not introduce bias because noisy forecasts only lead to measurement error in the left-hand side of the regression. Measurement error only makes the estimates less precise.

The right column shows the estimates obtained using local projections, our preferred method when the true shocks are not observed. Local projections only use the forecast errors, as given in Eq. (10). With local projections, measurement error leads to an attenuation bias. As discussed in Section 4.3, this issue is not unique to our method. The bias can be substantial if measurement error is large. Perhaps more surprisingly, measurement error does not make the estimates less precise. The reason is that measurement error is present in both the left- and right-hand side variables of the regression. While measurement error in the left-hand side variable unambiguously makes least squares estimates less precise, measurement error on the right-hand side introduces variation that is helpful for identifying the value of the biased coefficient. We stress that local projections correctly identify the IRF of the observed forecast errors. The issue is that with measurement error, the IRF of the observed forecast errors is no longer the object of economic interest.

Comparing the two columns, we observe that our method has reasonable statistical power. For example, when $R^2 = 0.75$, our method is able to detect the first three bias coefficients in 90% of Monte Carlo simulations, whereas with true shocks, we can detect the first five bias coefficients. In addition, the comparison is not entirely fair since in practice we are unlikely to observe all of the shocks. As a result, it may be more reasonable to compare our method when $R^2 = 0.75$ to that of using the true shocks when $R^2 = 0.25$ or so. To be concrete, to apply their method, Coibion and Gorodnichenko (2012) first estimate various shocks to inflation. Coibion and Gorodnichenko show that technology shocks explain the largest share of the variation in inflation of all the shocks that they consider. However, even technology shocks only explain around 25% of the variation (their Table 2). In contrast, local projections directly use all of the variation in

¹⁷ In their paper, Coibion and Gorodnichenko use a somewhat more parametric specification (see their Eq. (33)). In Appendix D, Coibion and Gorodnichenko consider a specification (their Eq. (D1)) that is equivalent to Eq. (9) and show that it yields similar results empirically.

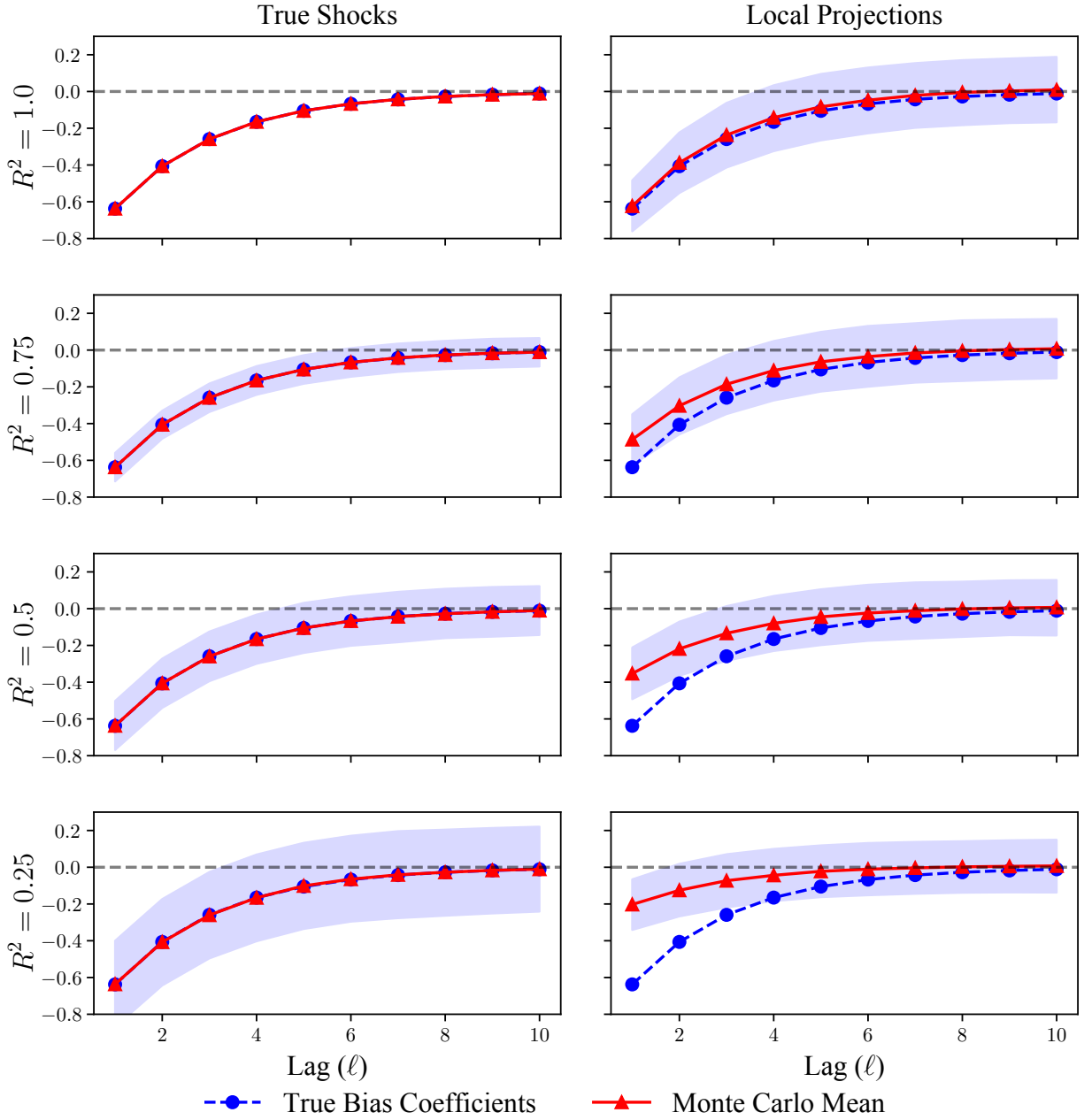


Figure 4: Baseline Monte Carlo results. The shaded areas give 90% of the Monte Carlo realizations (5th and 95th percentiles of the Monte Carlo realizations). 10,000 Monte Carlo replications are used for each figure. The top row has no measurement error; the lower rows have progressively more measurement error.

Left column: Bias coefficients estimated using the true shocks:

$$x_t - \mathbb{F}_{t-1}[x_t] = \alpha + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_{10} \varepsilon_{t-10} + u_t.$$

Right column: Bias coefficients estimated using local projections (not using the true shocks):

$$x_{t+s} - \mathbb{F}_{t+s-1}[x_{t+s}] = \alpha + \beta_1^{(s)} \{x_{t-1} - \mathbb{F}_{t-2}[x_{t-1}]\} + \dots + \beta_4^{(s)} \{x_{t-4} - \mathbb{F}_{t-5}[x_{t-4}]\} + u_t,$$

where $s = 0, 1, \dots, 9$.

the variable that is being forecast.

Nevertheless, the results in Figure 4 indicate that in some cases our method is unlikely to detect bias. For instance, consider expectations of stock market returns, as studied by Greenwood and Shleifer (2014), among others. If we assume that stock market returns are uncorrelated over time (as discussed in Section 2.3), then the estimates in Greenwood and Shleifer (2014, Table 3) suggest a value of 0.03 for the bias coefficient at the first yearly lag.¹⁸ Our method is unlikely to detect a bias of this magnitude. In the current simulation, our method fails to detect the bias coefficient at the fourth lag in many simulations, even when no measurement error is present. The magnitude of the bias coefficient at the fourth lag in the simulation is -0.165 , or roughly five times greater than the likely effect size in the stock market setting. This reasoning also suggests that there may be large efficiency gains from imposing additional structure when measuring overreaction in financial markets (such as imposing $\rho = 0$).

In addition to the baseline simulation, we investigate the robustness of our methodology to a number of variations. We consider two changes to the underlying x_t process. First, we suppose that x_t is not covariance stationary and follows a random walk (i.e., $x_t = x_{t-1} + \varepsilon_t$ and $x_0 = 0$), with all other parameters unchanged. Second, instead of normally distributed i.i.d. shocks, we suppose that the ε_t 's follow a GARCH(1, 1):

$$\begin{aligned}\varepsilon_{t+1} &= \sigma_{t+1} z_{t+1}, z_t \sim \mathcal{N}(0, 1) \\ \sigma_{t+1}^2 &= \omega + \alpha \varepsilon_t^2 + \beta \sigma_t^2\end{aligned}$$

To calibrate the process, we use the estimates for inflation reported by Capistrán and Timmerman (2009, Table 1): $\omega = 0.02$, $\alpha = 0.12$, and $\beta = 0.86$. These parameter values imply an unconditional variance of 1, which is very close to the unconditional variance of $\sigma_\varepsilon^2 = 1.005$ used in the baseline simulation. However, with a GARCH process, extreme realizations of ε_t are much more likely.

Finally, we examine the robustness of the baseline simulation by examining an alternative model of expectations. Existing experimental work has documented that subjects' forecasts are often well approximated by simple linear functions of past forecasts and

¹⁸ In Table 3, Greenwood and Shleifer regress expectations of stock market returns on realized stock market returns over the past year. For the case of Chief Financial Officers in the U.S. (column GH) who directly report a quantitative estimate of future stock returns, the estimated slope coefficient is 3.13 (s.e. = 1.25). Greenwood and Shleifer measure expectations in percentages (e.g., 5 for an expected return of 5%) whereas past stock returns are expressed as a fraction (e.g., 0.05 for a realized return of 5%), see their Table 1. Dividing the slope coefficient by 100 leads to the number reported in the text. The true value of the bias coefficient may, in fact, be lower if stock returns are somewhat autocorrelated, as discussed in Footnote 10.

past values of x_t .¹⁹ For that reason we also consider a simple model of extrapolative expectations:

$$\mathbb{F}_t[x_{t+1}] = x_t + \gamma(x_t - x_{t-1}). \quad (16)$$

In a recent experiment on inflation forecasting, [Pfafjar and Žakelj \(2014\)](#) document that many people use such forecasting rules. Based on their findings, we pick $\gamma = 0.50$.²⁰

Finally, we consider a Markov-switching model of expectations. We assume that expectations follow Eq. (16) but with one change: Expectations can now be in one of two states. They can either be trend seeking with $\gamma_H = 0.50$ or contrarian with $\gamma_L = -0.50$. The movement from the current γ to the value of γ in the next period, denoted by γ' , is governed by a Markov chain. The transition matrix is symmetric with

$$\mathbb{P}(\gamma' = \gamma_s | \gamma = \gamma_s) = 0.75$$

$$\mathbb{P}(\gamma' \neq \gamma_s | \gamma = \gamma_s) = 0.25$$

where $s \in \{L, H\}$. While the particular numbers are less empirically grounded, the specification captures the fact—well documented in experiments—that people often switch between different internal models of expectations.²¹ Prominent theoretical models that feature switching expectations include [Brock and Hommes \(1997\)](#) and [Barberis, Shleifer, and Vishny \(1998\)](#). Our formulation is particularly motivated by the latter paper.

The results are shown in Figure 5. For the case of GARCH errors and a random walk for x_t , the results are very similar to those reported in the baseline setting for $R^2 = 0.50$. The findings for extrapolative expectations are also similar. While both local projections and the estimation procedure that uses the true shocks is able to detect the first bias coefficient, local projections are subject to an attenuation bias from measurement error. With Markov-switching expectations, although the true bias coefficients are not zero, both methods often fail to detect bias. The reason is simple: Expectations switch between two opposite types of bias. If the econometrician fails to notice the Markov-switching structure of expectations and estimates the bias coefficients using the full sample, the resulting estimates yield a time average of the true coefficients, which is close to zero. The example highlights how ignoring changes in expectations

¹⁹ See, among many others, [Hey \(1994\)](#), [Heemeijer, Hommes, Sonnemans, and Tuinstra \(2009\)](#) and [Bao, Hommes, and Makarewicz \(2017\)](#).

²⁰ [Pfafjar and Žakelj \(2014, Table 5\)](#) find that for a plurality (but not a majority) of people, rational expectations cannot be rejected, and trend extrapolation is the second largest category of subjective expectations. While [Pfafjar and Žakelj \(2014\)](#) do not provide the average estimates for the trend extrapolation model, in Supplementary Material, they state that the estimated extrapolation coefficient lies between 0 and 1 in most cases. For that reason, we pick $\gamma = 0.50$.

²¹ See [Pfafjar and Žakelj \(2014\)](#) or [Assenza, Bao, Hommes, and Massaro \(2014\)](#).

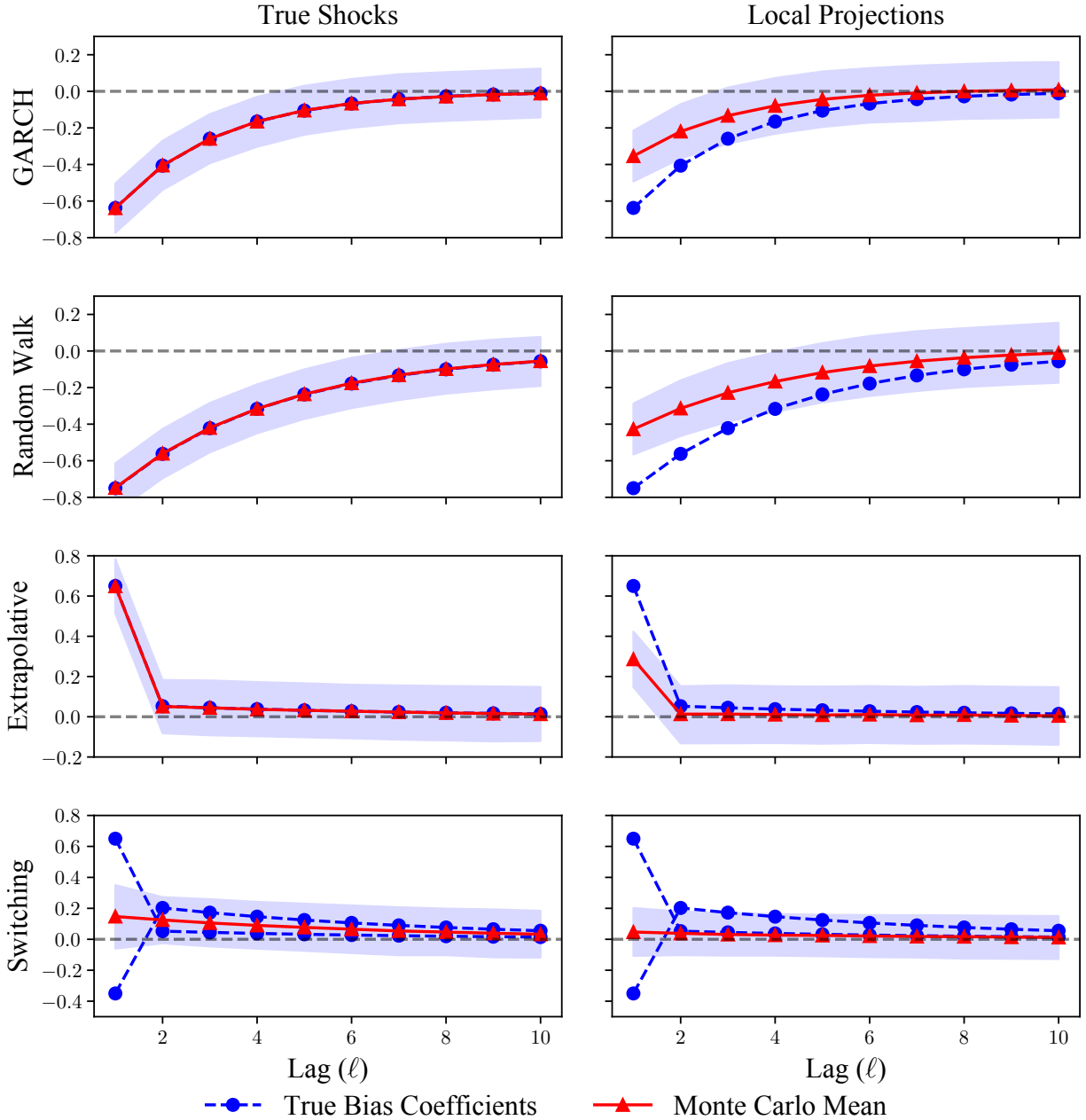


Figure 5: Additional Monte Carlo results. The shaded areas give 90% of the Monte Carlo realizations (5th and 95th percentiles of the Monte Carlo realizations). 10,000 Monte Carlo replications are used for each figure. The left column uses the true shocks to estimate bias coefficients; the right column uses local projections (without using the true shocks). The top row has no measurement error; the lower rows have progressively more measurement error. See Figure 4 for the exact regression specifications.

GARCH: GARCH(1, 1) distributed shocks. *Random Walk*: random walk for the underlying process ($\rho = 1$). *Extrapolative*: extrapolative expectations with $\mathbb{E}_t[x_{t+1}] = x_t + \gamma(x_t - x_{t-1})$ and $\gamma = 0.50$. *Switching*: Markov-switching model with extrapolative expectations in which $\gamma = 0.50$ or $\gamma = -0.50$ according to a Markov chain. The simulation with GARCH errors calibrates the value of σ_v so that $R^2 = 0.50$ using Eq. (15). All other simulations use the same level of measurement error $\sigma_v \approx 0.963$ as in Figure 4 for $R^2 = 0.50$.

| | N | \bar{x} | σ_x | ρ_x | \bar{e} | RMSE | ρ_e | R_{adj}^2 |
|-----------------------|-------|-----------|------------|----------|-----------|------|----------|--------------------|
| Inflation: Consensus | 196 | 3.45 | 2.59 | 0.83 | -0.05 | 1.40 | 0.40 | 0.20 |
| Inflation: Individual | 7,475 | 3.80 | 2.67 | 0.81 | 0.02 | 1.96 | 0.32 | 0.16 |

Table 1: Summary statistics. Root mean-squared error (RMSE) is calculated as $\sqrt{\frac{1}{N} \sum e_t^2}$ where e_t is the forecast error. Persistence ρ_z is measured by the estimate of b in the regression $z_t = a + bz_{t-1} + v_t$. R_{adj}^2 is the adjusted R -squared in the regression of forecast errors on the past four forecast errors $e_{t-1}, e_{t-2}, \dots, e_{t-4}$.

may significantly bias estimates.

6 Application: Inflation Expectations

We now turn to an empirical application of our method.

The application uses inflation forecasts from the Survey of Professional Forecasters (SPF). The SPF is currently run by the Federal Reserve Bank of Philadelphia. Each quarter, participants of the survey forecast a number of macroeconomic and financial variables. The names of the individual participants in the survey are not publicly known, limiting the scope for strategic considerations. The participants are professional forecasters; see [Croushore \(1993\)](#) for further discussion.

The dataset has been used extensively in prior work, with recent studies including [Carroll \(2003\)](#), [Mankiw, Reis, and Wolfers \(2003\)](#), [Coibion and Gorodnichenko \(2012\)](#), and [Coibion and Gorodnichenko \(2015\)](#).²² As a result, the dataset provides a natural testing ground for our method. To streamline the discussion, we focus on the findings and explain how we construct the dataset in [Appendix C](#).

6.1 Bias Coefficients

We study one-quarter ahead GDP deflator inflation forecasts. Summary statistics of the dataset are provided in [Table 1](#).²³ Both consensus (median) and individual-level forecasts are considered. We construct the consensus forecasts from the individual forecasts available on the SPF website. We use median forecasts for the consensus to be consistent with prior work, but the median and mean forecasts are very similar. To avoid bias stemming from data revisions, we use real-time data to measure realized inflation.

²² The Philadelphia Fed provides an [extensive list](#) of studies that have used this dataset.

²³ The number of participants in the SPF has not been constant over time. As a result, estimates using the consensus- and individual-level datasets implicitly weight the data somewhat differently. The individual-level dataset implies a somewhat higher weight on observations coming from the earlier part of the sample. That is the reason why, for instance, the estimates of mean inflation differ across these two datasets.

The results are shown in Figure 6. The top panel plots the bias coefficients estimated using consensus forecasts. To calculate the bias coefficients, we first estimate the IRF of forecast errors using local projections, as in Eq. (10). Then, the estimated IRF is multiplied by (-1) to obtain the bias coefficients.²⁴ To be conservative, we show both the 65% and 95% confidence intervals calculated using Newey-West standard errors. The resulting bias coefficients are *aggregate* bias coefficients in the sense of Section 2.1.2. The estimated bias coefficients measure the total responsiveness to shocks and do not disentangle between the reaction to particular shocks (e.g., we cannot say whether agents underreact to oil price shocks or monetary policy shocks).

We observe statistically significant negative bias coefficients for lag 1 ($p = 0.005$),²⁵ lag 3 ($p = 0.022$), and lag 4 ($p < 0.001$). The evidence suggests that participants in the SPF underreact to information that arrived up to one year ago. In Appendix D, we show that virtually identical results obtain—with somewhat smaller standard errors—if instead of local projections, we use maximum likelihood to fit a high-order moving average model (Figure 7).

The magnitude of underreaction is substantial. The point estimates indicate that a positive 1σ shock to inflation in the current quarter leads the forecasters to underpredict inflation by roughly 0.30σ four quarters from now.

The bottom panel of Figure 6 performs the same exercise using the individual-level forecasts. We estimate a panel-data equivalent of Eq. (10), including forecaster fixed effects. To account for the fact that the respondents are all forecasting the same variable, and the forecast errors may be correlated over time for a given respondent, we cluster the standard errors by both individual forecaster and quarter, see Thompson (2011) and Colin Cameron, Gelbach, and Miller (2011).

The pattern of the estimated bias coefficients is very similar to that obtained using the consensus forecasts. However, the bias coefficient at lag 3 is no longer statistically significant at the 5% level ($p = 0.096$). Overall, the point estimates from the individual-level data tend to be smaller in absolute value. However, they fall well within the confidence intervals obtained using consensus forecasts. Of course, the result is not unexpected given the theoretical aggregation result in Section 4.2.

²⁴ As shown in Appendix D (Figure 8), the IRF of inflation is positive at all the relevant lags, so that $\text{sgn}(\alpha_\ell) = 1$.

²⁵ At the time of responding to the survey, participants know the advance estimate of inflation in the previous quarter but not inflation in the current quarter (Federal Reserve Bank of Philadelphia, 2017, p. 21). As a result, interpretation of the bias coefficient in the first lag requires some care. On the one hand, forecasters have access to various real-time information on prices. On the other hand, they do not yet know the official number for inflation in the current quarter. In that sense, the one-step ahead forecast may really be a two-step ahead forecast. If that is the case, a non-zero bias coefficient at the first lag should not be interpreted as bias. See Keane and Runkle (1990) for further discussion.

The fact that we obtain very similar results using both consensus- and individual-level forecasts may seem surprising given the recent empirical findings by [Bordalo, Gennaioli, Ma, and Shleifer \(2018a\)](#) and [Broer and Kohlhas \(2018\)](#). These authors use SPF data to document that forecast errors are positively correlated with past forecast revisions at the consensus level (as in [Coibion and Gorodnichenko, 2015](#)), but the two variables are negatively correlated at the individual level. These findings suggest underreaction at the consensus level but overreaction at the level of individual forecasters. Our preferred interpretation is somewhat different: The findings are consistent with forecasters overreacting to some information while *simultaneously* underreacting to other information. For example, forecasters may be overreacting to some private information, proxied by their own forecast revision, but underreacting to publicly available information (see [Section A.6](#) for a formalization). Alternatively, these findings may indicate measurement error, as discussed in [Section 4.3](#). Overall, our own empirical findings suggest that the dominant force is underreaction.

6.2 Calibration Exercise

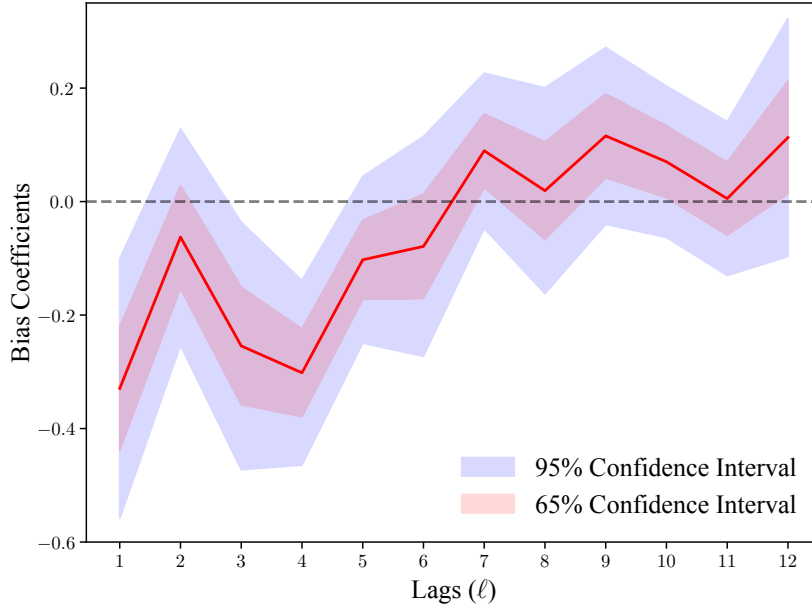
The estimated bias coefficients can be used to guide theory. To illustrate this point, we perform a simple calibration exercise. For a number of models, we choose their parameters to fit the estimated bias coefficients as closely as possible. Specifically, for each model of expectations, we choose its parameters θ to minimize the sum of squared residuals

$$\text{SSR} = \sum_{\ell=1}^{12} [\hat{b}_{\ell} - b_{\ell}(\theta)]^2,$$

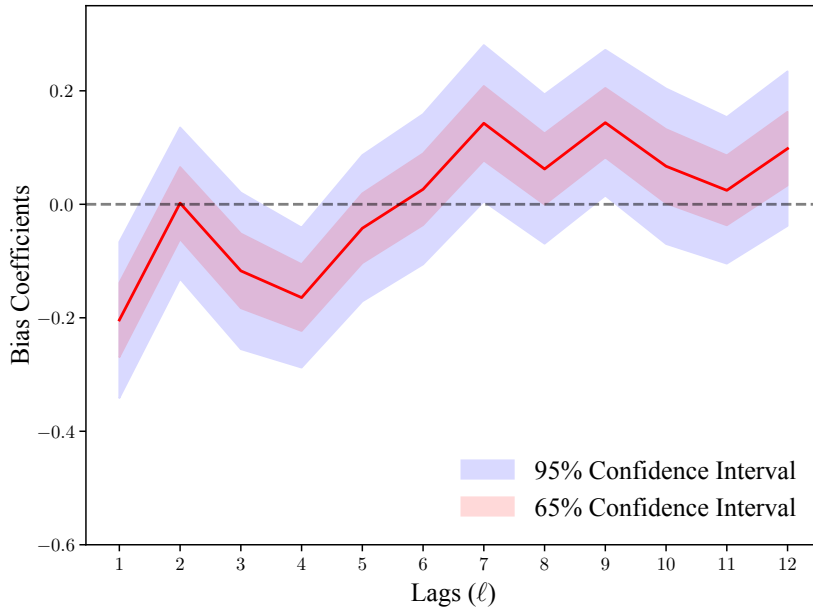
where \hat{b}_{ℓ} is the empirically estimated bias coefficient, and $b_{\ell}(\theta)$ is the theoretically predicted bias coefficient. To obtain theoretical predictions, we assume that the true inflation process is an AR(1). Estimating the persistence parameter using least squares yields an estimate of $\hat{\rho} = 0.83$ (see [Table 1](#)).

The results from this exercise are shown in [Table 2](#). Since the models all have a single parameter (except for rational expectations which have no free parameters), we do not adjust for model complexity. The model that best fits the data is a simple model in which forecasters think that the true persistence of inflation is smaller than it actually is (see [Section A.3](#)). The perceived level of persistence that provides the best fit is 0.61. This number is roughly 25% lower than the estimated persistence of inflation. As discussed by [Gabaix \(2017a, pp. 14–15\)](#), limited attention can naturally lead to a misperception of persistence.

The sticky information model also does well, with only a slightly worse fit than the



(a) Inflation: Consensus-Level Estimates



(b) Inflation: Individual-Level Estimates

Figure 6: Bias coefficients for one-quarter ahead inflation forecasts. *Top panel*: estimates using consensus (median) forecasts; Newey-West standard errors with four lags are used to calculate the confidence intervals. *Bottom panel*: estimates using individual-level data (controlling for forecaster fixed effects); standard errors are clustered by both forecaster and quarter. Both sets of estimates are obtained by first using local projections (with $K = 4$) to estimate the impulse response function of the forecast errors:

$$x_{t+s} - \mathbb{F}_{t+s-1}[x_{t+s}] = \alpha + \beta_1^{(s)} \{x_{t-1} - \mathbb{F}_{t-2}[x_{t-1}]\} + \cdots + \beta_4^{(s)} \{x_{t-4} - \mathbb{F}_{t-5}[x_{t-4}]\} + u_t,$$

where $s = 0, 1, \dots, 11$. The bias coefficients are then estimated by $\hat{b}_\ell = -\hat{\beta}_1^{(\ell-1)}$.

| Model | Parameter | SSR |
|-----------------|-----------|------|
| Misperception | 0.61 | 0.16 |
| Sticky | 0.51 | 0.18 |
| Adjustment Cost | 0.43 | 0.25 |
| Rational | NA | 0.32 |
| Diagnostic | 0.00 | 0.32 |
| Adaptive | 0.17 | 0.36 |
| Extrapolative | -0.31 | 0.55 |

Table 2: Calibration exercise for matching the empirically estimated bias coefficients. The sum of squared residuals (SSR) is calculated as $\sum_{\ell=1}^{12} [\hat{b}_\ell - b_\ell(\theta^*)]^2$ where \hat{b}_ℓ is the empirically estimated bias coefficient, and $b_\ell(\theta^*)$ is the theoretically predicted bias coefficient; θ^* denotes the parameter value that minimizes the sum of squared residuals. The estimated process for inflation is $x_t = 0.83x_{t-1} + \varepsilon_t$. The precise descriptions of the models are given in Section 3 and Appendix A.

misperception model. The estimated information stickiness parameter for this model is $\hat{\lambda} = 0.51$. The estimate is very close to that reported by [Coibion and Gorodnichenko \(2015\)](#) who find $\hat{\lambda} = 0.54$ (s.e. = 0.10).²⁶ While [Coibion and Gorodnichenko \(2015\)](#) also use the SPF data, they consider one-year ahead forecasts, and their methodology estimates the level of stickiness by regressing forecast errors on past forecast revisions. The fact that we get a very similar number using a completely different methodology is reassuring. In the context of the sticky information model, our estimate implies that forecasters in the SPF update their information sets roughly twice a year on average.

The model with adjustment costs performs better than simple rational expectations. However, in the current setting it is not exactly clear what the adjustment cost may represent, given that the identities of the participants are not publicly known. Diagnostic expectations in this example do exactly as well as rational expectations, since they cannot generate underreaction. Finally, mechanical adaptive and extrapolative expectations perform worse than rational expectations. The finding is consistent with previous research that has documented that participants in the SPF are quite accurate.²⁷ It is therefore not surprising that their behavior is not very well described by mechanical models of expectations. The fact that extrapolative expectations perform especially poorly is interesting in light of the fact that extrapolative models explain inflation expectations well in laboratory experiments (e.g., [Pfajfar and Žakelj, 2014](#)).

²⁶ We use the Delta Method to calculate the standard error for $\hat{\lambda}$ from the estimates provided by [Coibion and Gorodnichenko](#), i.e., $\text{s.e.}(\hat{\lambda}) = \text{s.e.}(\hat{\beta})/(1 + \hat{\beta})^2$.

²⁷ See, for example, [Croushore \(2010\)](#). [Croushore](#) documents that the median forecast in the SPF performs better than simple time series models of inflation (Table 5), and that it is difficult to adjust forecasts for biases observed in the past to obtain higher forecasting accuracy (Table 4).

7 Conclusions

We have developed a new framework for measuring biases in expectation formation. The framework allows for a fairly unified treatment of existing models, and it provides a simple way to measure biases empirically. The estimated bias coefficients can be used to distinguish between different models of expectations. Since our empirical method is based on local projections, it is easy to accommodate various empirical scenarios, including panel data and clustered error terms. The flexibility of local projections can also be brought to bear on questions that may be difficult to address otherwise, such as whether expectations exhibit non-linearities or state dependence.

The proposed method imposes fewer assumptions than existing approaches. The additional generality can come at the cost of lower statistical precision. When considering using our method, researchers should weigh the costs of imposing additional assumptions—which may be violated in practice—against the benefits of increased statistical power. Additional structure is likely to be especially valuable when the dataset has a small time series dimension.

The notion of bias used in this paper is a statistical one. There may be multiple reasons for why expectations are biased according to this definition, including psychological (e.g., belief persistence) as well as non-psychological ones (e.g., informational frictions). A related challenge is that different models of expectations can yield similar predictions for bias coefficients. Predictions on additional moments of the data can help to distinguish between competing explanations in these situations.²⁸

Given the methodological nature of the present work, we have illustrated the method with a single yet prominent application: inflation expectations. Many new datasets on subjective expectations have become available in recent years, thanks to numerous data collection efforts. While much remains to be done, new data sources will no doubt shed light on how expectations are formed. Our method may prove useful in this endeavor.

²⁸ For instance, the models of sticky information and noisy information imply identical values for the bias coefficients when the underlying process is an AR(1). However, the two models make different predictions on how disagreement among forecasters changes in response to shocks (Coibion and Gorodnichenko, 2012).

References

- Abarbanell, J. S., Bernard, V. L., 1992. Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *The Journal of Finance* 47 (3), 1181–1207.
- Anderson, O. D., 1975. On a lemma associated with Box, Jenkins and Granger. *Journal of Econometrics* 3 (2), 151–156.
- Andreassen, P. B., 1987. On the social psychology of the stock market: Aggregate attributional effects and the regressiveness of prediction. *Journal of Personality and Social Psychology* 53 (3), 490–496.
- Ansley, C. F., Spivey, W. A., Wroblewski, W. J., 1977. On the structure of moving average processes. *Journal of Econometrics* 6 (1), 121–134.
- Assenza, T., Bao, T., Hommes, C., Massaro, D., 2014. Experiments on expectations in macroeconomics and finance. In: Duffy, J. (Ed.), *Experiments in Macroeconomics (Research in Experimental Economics)*. Vol. 17. Emerald Group Publishing Limited, Bingley, pp. 11–70.
- Augenblick, N., Lazarus, E., 2018. Restrictions on asset-price movements under rational expectations: Theory and evidence. Working Paper.
- Augenblick, N., Rabin, M., 2018. Belief movement, uncertainty reduction, and rational updating. Working Paper.
- Bao, T., Hommes, C., Makarewicz, T., 2017. Bubble formation and (in)efficient markets in learning-to-forecast and optimise experiments. *The Economic Journal* 127, 581–609.
- Barberis, N., Greenwood, R., Jin, L., Shleifer, A., 2015. X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics* 115 (1), 1–24.
- Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. *Journal of Financial Economics* 49 (3), 307–343.
- Blackwell, D., Dubins, L., 1962. Merging of opinions with increasing information. *The Annals of Mathematical Statistics* 33 (3), 882–886.
- Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2018a. Overreaction in macroeconomic expectations. Working Paper.

- Bordalo, P., Gennaioli, N., Shleifer, A., 2018b. Diagnostic expectations and credit cycles. *The Journal of Finance* 73 (1), 199–227.
- Bouchaud, J.-P., Krüger, P., Landier, A., Thesmar, D., 2018. Sticky expectations and the profitability anomaly. *The Journal of Finance*, forthcoming.
- Brock, W. A., Hommes, C., 1997. A rational route to randomness. *Econometrica* 65 (5), 1059–1095.
- Broer, T., Kohlhas, A. N., 2018. Forecaster (mis-)behavior. CEPR Discussion Paper No. 12898.
- Cagan, P., 1956. The monetary dynamics of hyperinflation. In: Friedman, M. (Ed.), *Studies in the Quantity Theory of Money*. University of Chicago Press, Chicago, pp. 25–117.
- Capistrán, C., Timmerman, A., 2009. Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking* 41 (2), 365–396.
- Carroll, C. D., 2003. Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics* 118 (1), 269–298.
- Chapman, L. J., Chapman, J., 1982. Test results are what you think they are. In: Kahneman, D., Slovic, P., Tversky, A. (Eds.), *Judgment under Uncertainty*. Cambridge University Press, Cambridge, Ch. 17.
- Cochrane, J., 2005. *Asset Pricing: Revised Edition*. Princeton University Press, Princeton.
- Coibion, O., Gorodnichenko, Y., 2012. What can survey forecasts tell us about information rigidities? *Journal of Political Economy* 120 (1), 116–159.
- Coibion, O., Gorodnichenko, Y., 2015. Information rigidity and the expectations formation process: A simple framework and new facts. *The American Economic Review* 105 (8), 2644–2678.
- Coibion, O., Gorodnichenko, Y., Kamdar, R., 2018. The formation of expectations, inflation and the Phillips Curve. *Journal of Economic Literature*, forthcoming.
- Colin Cameron, A., Gelbach, J. B., Miller, D. L., 2011. Robust inference with multiway clustering. *Journal of Business and Economic Statistics* 29 (2), 238–249.

- Croushore, D., 2010. An evaluation of inflation forecasts from surveys using real-time data. *The B.E. Journal of Macroeconomics* 10 (1).
- Croushore, D. D., 1993. Introducing: The Survey of Professional Forecasters. *Business Review – Federal Reserve Bank of Philadelphia* 3 (December), 3–15.
- Cutler, D. M., Poterba, J. M., Summers, L. H., 1990a. Speculative dynamics. *Review of Economic Studies* 58 (3), 529–546.
- Cutler, D. M., Poterba, J. M., Summers, L. H., 1990b. Speculative dynamics and the role of feedback traders. *The American Economic Review: Papers & Proceedings* 80 (2), 63–68.
- Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. *The Journal of Finance* 53 (6), 1839–1885.
- De Bondt, W. F. M., Thaler, R. H., 1990. Do security analysts overreact? *The American Economic Review: Papers & Proceedings* 80 (2), 52–57.
- DeLong, B. J., Shleifer, A., Summers, L. H., Waldmann, R. J., 1990. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance* 45 (2), 379–395.
- Diebold, F. X., Lopez, J. A., 1996. Forecast evaluation and combination. In: Maddala, G., Rao, C. (Eds.), *Handbook of Statistics (Statistical Methods in Finance)*. Vol. 14. Elsevier, Amsterdam, Ch. 8, pp. 241–268.
- Dominitz, J., Manski, C. F., 2011. Measuring and interpreting expectations of equity returns. *Journal of Applied Econometrics* 26, 352–370.
- Durbin, J., Koopman, S. J., 2012. *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.
- Edwards, W., 1968. Conservatism in human information processing. In: Kleinmuntz, B. (Ed.), *Formal Representation of Human Judgement*. John Wiley and Sons, New York.
- Evans, G. W., Honkapohja, S., 2001. *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton.
- Federal Reserve Bank of Philadelphia, 2017. *Survey of Professional Forecasters: Documentation*. [Link to website](#).

- Fisher, F., 1962. *A Priori Information and Time Series Analysis: Essays in Economic Theory and Measurement*. Amsterdam: North-Holland.
- Fuster, A., Laibson, D., Mendel, B., 2010. Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives* 24 (4), 67–84.
- Gabaix, X., 2017a. Behavioral inattention. Working Paper, Harvard University.
- Gabaix, X., 2017b. Behavioral macroeconomics via sparse dynamic programming. Working Paper, Harvard University.
- Gabaix, X., 2018. A behavioral New Keynesian model. Working Paper, Harvard University.
- Gennaioli, N., Shleifer, A., 2010. What comes to mind. *The Quarterly Journal of Economics* 125 (4), 1399–1433.
- Gilovich, T., Vallone, R., Tversky, A., 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17, 295–314.
- Goodwin, R. M., 1947. Dynamical coupling with especial reference to markets having production lags. *Econometrica* 15 (3), 181–204.
- Graham, J. R., Harvey, C. R., 2001. Expectations of equity risk premia, volatility and asymmetry from a corporate finance perspective. NBER Working Paper No. 8678.
- Granger, C. W. J., 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14 (2), 227–238.
- Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. *Review of Financial Studies* 27 (3), 714–746.
- Griffin, D., Tversky, A., 1992. The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24, 411–435.
- Hamilton, J. D., 1994. *Time Series Analysis*. Princeton University Press, Princeton.
- Heemeijer, P., Hommes, C., Sonnemans, J., Tuinstra, J., 2009. Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. *Journal of Economic Dynamics & Control* 33, 1052–1072.
- Hey, J. D., 1994. Expectations formation: Rational or adaptive or ...? *Journal of Economic Behavior and Organization* 25 (3), 329–349.

- Jordá, Ò., 2005. Estimation and inference of impulse responses by local projections. *The American Economic Review* 95 (1), 161–182.
- Kahneman, D., Tversky, A., 1973. On the psychology of prediction. *Psychological Review* 80 (4), 237–251.
- Keane, M. P., Runkle, D. E., 1990. Testing the rationality of price forecasts: New evidence from panel data. *The American Economic Review* 80 (4), 714–735.
- Kydland, F. E., Prescott, E. C., 1982. Time to build and aggregate fluctuations: Some new evidence. *Econometrica* 50 (6), 1345–1370.
- Langer, E. J., 1975. The illusion of control. *Journal of Personality and Social Psychology* 32, 311–328.
- Ma, Y., Landier, A., Thesmar, D., 2017. New experimental evidence on expectations formation. Working Paper.
- Mankiw, G. N., Reis, R., 2002. Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics* 117 (4), 1295–1328.
- Mankiw, N. G., Reis, R. A., Wolfers, J., 2003. Disagreement about inflation expectations. *NBER Macroeconomics Annual* 18, 209–270.
- Manski, C. F., 2004. Measuring expectations. *Econometrica* 72 (5), 1329–1376.
- Manski, C. F., 2018. Survey measurement of probabilistic macroeconomic expectations: Progress and promise. *NBER Macroeconomics Annual* 32, forthcoming.
- Marimon, R., 1997. Learning from learning in economics. In: Kreps, D. M., Wallis, K. F. (Eds.), *Advances in Economics and Econometrics: Theory and Applications*. Vol. 1. Cambridge University Press, Cambridge, Ch. 9, pp. 278–315.
- Miller, J. B., Sanjurjo, A., 2018. Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica*, forthcoming.
- Mincer, J. A., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. A. (Ed.), *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. National Bureau of Economic Research, Ch. 1, pp. 3–46.
- Muth, J. F., 1961. Rational expectations and the theory of price movements. *Econometrica* 29 (3), 315–335.

- Nerlove, M., 1958. Adaptive expectations and cobweb phenomena. *The Quarterly Journal of Economics* 72 (2), 227–240.
- Nickerson, R. S., 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2 (2), 175–220.
- Nisbett, R., Ross, L., 1980. *Human Inference: Strategies and Shortcomings of Social Judgement*. Prentice Hall, Englewood Cliffs.
- Odean, T., 1998. Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance* 53 (6), 1887–1934.
- Patton, A. J., Timmermann, A., 2007. Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics* 140 (2), 884–918.
- Pesaran, M. H., Weale, M., 2006. Survey expectations. In: Elliott, G., Granger, C. W. J., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*. Vol. 1. Elsevier, Amsterdam, Ch. 14, pp. 715–776.
- Pfajfar, D., Žakelj, B., 2014. Experimental evidence on inflation expectation formation. *Journal of Economic Dynamics & Control* 44, 147–168.
- Ramey, V. A., 2016. Macroeconomic shocks and their propagation. In: Taylor, J. B., Uhlig, H. (Eds.), *Handbook of Macroeconomics*. Vol. 2. Elsevier, Amsterdam, Ch. 2, pp. 71–162.
- Reis, R., 2006. Inattentive producers. *Review of Economic Studies* 73 (3), 793–821.
- Schmalensee, R., 1976. An experimental study of expectation formation. *Econometrica* 44 (1), 17–41.
- Sims, C. A., 2003. Implications of rational inattention. *Journal of Monetary Economics* 50 (3), 665–690.
- Staudenmayer, J., Buonaccorsi, J. P., 2005. Measurement error in linear autoregressive models. *Journal of the American Statistical Association* 100, 841–852.
- Thompson, S. B., 2011. Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics* 99 (1), 1–10.
- Vissing-Jorgensen, A., 2004. Perspectives on behavioral finance: Does ‘irrationality’ disappear with wealth? Evidence from expectations and actions. *NBER Macroeconomics Annual* 18, 139–208.

Woodford, M., 2003. Imperfect common knowledge and the effects of monetary policy. In: Aghion, P., Frydman, R., Stiglitz, J., Woodford, M. (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*. Princeton University Press, Princeton, pp. 25–58.

Appendix A Mapping Existing Models

A.1 Adaptive Expectations

Suppose that the agent has adaptive expectations as in [Cagan \(1956\)](#) and [Nerlove \(1958\)](#):

$$\mathbb{F}_t[x_{t+1}] = \mathbb{F}_{t-1}[x_t] + \lambda\{x_t - \mathbb{F}_{t-1}[x_t]\}.$$

Iterating, we have that

$$\begin{aligned} \mathbb{F}_t[x_{t+1}] &= \lambda \sum_{\ell=0}^{+\infty} (1-\lambda)^\ell x_{t-\ell} = \lambda \sum_{\ell=0}^{+\infty} \sum_{j=0}^{+\infty} (1-\lambda)^\ell \rho^j \varepsilon_{t-\ell-j} \\ &= \lambda \sum_{\ell=0}^{+\infty} \left(\frac{(1-\lambda)^{\ell+1} - \rho^{\ell+1}}{1-\lambda-\rho} \right) \varepsilon_{t-\ell}. \end{aligned}$$

Hence, we obtain

$$\begin{aligned} a_\ell &= \lambda \left[\frac{(1-\lambda)^\ell - \rho^\ell}{1-\lambda-\rho} \right] \\ b_\ell &= \text{sgn}(\rho^\ell) \left[\frac{\lambda(1-\lambda)^\ell - (1-\rho)\rho^\ell}{1-\lambda-\rho} \right] \end{aligned}$$

A.2 Noisy Information

We now analyze a model in which agents are rational and understand the structure of the model but do not observe the underlying state perfectly. Models of this type include the rational inattention model of [Sims \(2003\)](#) and the imperfect information model studied by [Woodford \(2003\)](#). The section follows [Coibion and Gorodnichenko \(2015, pp. 2649–2650\)](#) closely.

Suppose the true process for x_t is an AR(1) but each agent i only observes a noisy signal y_{it} of x_t :

$$y_{it} = x_t + \omega_{it}.$$

Here, ω_{it} is a normally distributed mean-zero noise term which is i.i.d. across time and agents. The Kalman filter equations then imply (see Eq. 8 in [Coibion and Gorodnichenko \(2015\)](#)) that

$$\mathbb{F}_{it}[x_{t+1}] = \rho \{G y_{it} + (1-G) \mathbb{F}_{i,t-1}[x_t]\},$$

where $G \in [0, 1]$ is the Kalman gain.

Assuming that there is a continuum of agents and making the usual Law of Large Numbers assumption, we can integrate over agents to find

$$\mathbb{F}_t[x_{t+1}] = G\rho x_t + (1 - G)\rho \mathbb{F}_{t-1}[x_t], \quad (17)$$

with $\mathbb{F}_t[x_{t+1}]$ denoting the average forecast across agents.

Now write $\mathbb{F}_t[x_{t+1}] = \sum_{\ell=0}^{+\infty} a_{\ell+1} \varepsilon_{t-\ell}$ for some unknown coefficients to be determined. From Eq. (17), we have that

$$a_1 \varepsilon_t + \sum_{\ell=1}^{+\infty} [a_{\ell+1} - (1 - G)\rho a_\ell] \varepsilon_{t-\ell} = G\rho \varepsilon_t + G \sum_{\ell=1}^{+\infty} \rho^{\ell+1} \varepsilon_{t-\ell}.$$

Matching coefficients and solving the resulting difference equation, we find that

$$\begin{aligned} a_\ell &= \rho^\ell - [(1 - G)\rho]^\ell \\ b_\ell &= -\operatorname{sgn}(\rho^\ell) [(1 - G)\rho]^\ell \end{aligned}$$

Hence, as long as the signal is not perfectly revealing of the state ($G \neq 1$), the noisy information model predicts underreaction. In fact, the model predicts identical bias coefficients as the sticky information model when $\lambda = 1 - G$.

A.3 Misperceived Law of Motion

Suppose that the agent misperceives the true persistence of the process and makes forecasts as

$$\mathbb{F}_t[x_{t+1}] = \hat{\rho} x_t, \hat{\rho} \in (-1, 1),$$

with $\hat{\rho}$ potentially different from ρ . Examples of models with misperceived laws of motion abound in the literature, with two prominent cases given by [Barberis, Shleifer, and Vishny \(1998\)](#) and [Fuster, Laibson, and Mendel \(2010\)](#); see also [Gabaix \(2017a\)](#), pp. 14–15). In the present case,

$$\mathbb{F}_t[x_{t+1}] = \hat{\rho} \sum_{\ell=0}^{+\infty} \rho^\ell \varepsilon_{t-\ell}$$

and therefore

$$\begin{aligned} a_\ell &= \hat{\rho} \rho^{\ell-1} \\ b_\ell &= \operatorname{sgn}(\rho^\ell) [\hat{\rho} \rho^{\ell-1} - \rho^\ell] \end{aligned}$$

When $\rho \neq 0$, we can write

$$b_\ell = \text{sgn}(\rho^\ell) \rho^\ell \left(\frac{\hat{\rho} - \rho}{\rho} \right).$$

If $\rho > 0$, the agent overreacts to news whenever $\hat{\rho} > \rho$ and underreacts otherwise.

A.4 Extrapolative Expectations

We now consider pure extrapolative expectations

$$\mathbb{F}_t[x_{t+1}] = x_t + \gamma(x_t - x_{t-1}),$$

as in [Goodwin \(1947, p. 191\)](#). The parameter γ could be either positive or negative, with a positive γ representing extrapolation or trend following, while a negative γ could capture contrarian expectations.

Substituting in the expression for x_t , we calculate that

$$\mathbb{F}_t[x_{t+1}] = (1 + \gamma)\varepsilon_t + \sum_{\ell=1}^{+\infty} \{(1 + \gamma)\rho^\ell - \gamma\rho^{\ell-1}\} \varepsilon_{t-\ell},$$

and so we find that

$$a_\ell = \begin{cases} 1 + \gamma & \text{if } \ell = 1 \\ (1 + \gamma)\rho^{\ell-1} - \gamma\rho^{\ell-2} & \text{if } \ell \geq 2 \end{cases},$$

and

$$b_\ell = \begin{cases} \text{sgn}(\rho)(1 + \gamma - \rho) & \text{if } \ell = 1 \\ \text{sgn}(\rho^\ell) \{(1 + \gamma)\rho^{\ell-1} - \gamma\rho^{\ell-2} - \rho^\ell\} & \text{if } \ell \geq 2 \end{cases}$$

Suppose expectations are of the trend-following type ($\gamma > 0$). Then, if x_t is positively autocorrelated, the agent always overreacts to current news ($b_1 > 0$). However, extrapolative expectations may well lead to underreaction to past news.

A.5 Learning

Many models of expectations study agents that are learning the true data-generating process over time. These models naturally lead to time-varying expectations. As a result, models with learning are generally not nested by our time-invariant formulation.

We could add a time t subscript to our specification without any theoretical difficulty. The problem is a practical one: Without assuming some constancy of parameters or imposing some other structure on expectations, measuring bias coefficients empirically does not seem feasible.

Models of learning often have the property that agents eventually learn the true model. According to the Blackwell-Dubins Theorem (Blackwell and Dubins, 1962), that is the case if agents are Bayesian, and—very loosely speaking—place a positive prior probability on the true model.²⁹ Another example is if the agents do not know the true data-generating process but estimate correctly specified regressions, as in many models of *adaptive learning* (Evans and Honkapohja, 2001). In these circumstances, a natural empirical approach is to split the sample into early and late periods and estimate the bias coefficients separately in the two subsamples. The prediction of learning is that the estimated bias coefficients should be closer to zero in the later subsample.

Finally, in some models, the agents know the true model but are not able to observe the true state. Such models involve learning about the true state (although not the parameters of the model). For example, the variable being forecast may consist of a transitory and a permanent component, but the agent may not be able to distinguish between the two. An example is given by the seminal paper by Kydland and Prescott (1982) who assume that the technology shock consists of a permanent and transitory component (see their equation 3.7). If the model can be written as a linear state space model with normally distributed disturbances, then a key result in the theory of Kalman filtering states that the forecast errors are independent.³⁰ In these cases, the predicted bias coefficients are zero.

A.6 Overconfidence

Finally, we discuss a stylized model of overconfidence along the lines of Daniel, Hirshleifer, and Subrahmanyam (1998) and Odean (1998), among others. Suppose that expectations follow

$$\mathbb{F}_t[x_{t+1}] = s_t, \quad (18)$$

where s_t is white noise (and independent of ε_t). We think of this specification as capturing an extreme form of overconfidence. In this interpretation, s_t is a private signal that is uncorrelated with actual news.

According to the definition in Section 2.1.2, expectations simultaneously overreact to the private signal (s_t) and underreact to the true shock (ε_t). Overall, expectations

²⁹ Marimon (1997, pp. 285–286) discusses this result and its implications in economics.

³⁰ See, for example, Durbin and Koopman (2012, p. 69).

exhibit underreaction. To see the latter point, note that we can write $(1 - \rho L)e_t = \varepsilon_t - s_{t-1} + \rho s_{t-2}$. The right-hand side is a sum of an MA(1) process and white noise, and therefore also an MA(1) process (see, e.g., [Hamilton, 1994](#), pp. 102–105). Write the right-hand side as $\xi_t + \theta \xi_{t-1}$ for some θ and a white noise series ξ_t . Then, the Wold Representation of e_t is given by³¹

$$e_t = \xi_t + (\rho + \theta) \sum_{\ell=1}^{+\infty} \rho^{\ell-1} \xi_{t-\ell}.$$

One can show that $\rho + \theta > 0$, and therefore expectations exhibit overall underreaction.

As these calculations make clear, the present overconfidence model is isomorphic to the model with a misperceived law of motion (Section [A.3](#) with $\hat{\rho} = 0$) and measurement error. Empirically, it may be challenging to distinguish between these two possibilities.

³¹ The calculation is exactly the same as the one used to calculate the bias coefficients in the presence of measurement error in [Appendix B.1](#), and the expression for θ is given in [Eq. \(20\)](#) with $(\lambda\rho)$ replaced with ρ .

Appendix B Measurement Error

B.1 Formulas for Sticky and Diagnostic Expectations

This appendix provides explicit formulas for the attenuation bias caused by measurement error for two models of expectations (sticky information and diagnostic expectations). As in Section 3, x_t follows a stationary AR(1) process.

First, suppose that the true expectations are generated by the sticky information model of [Mankiw and Reis \(2002\)](#), implying that the true forecast errors follow

$$e_t = \lambda\rho e_{t-1} + \varepsilon_t.$$

However, instead of observing the true forecast $\mathbb{F}_t[x_{t+1}]$, we can only observe

$$\mathbb{F}_t^*[x_{t+1}] = \mathbb{F}_t[x_{t+1}] + v_t,$$

where v_t is white noise measurement error with variance σ_v^2 (and independent of ε_t). The observed forecast error is then equal to $e_t^* = e_t - v_{t-1}$. Now write

$$e_t^* - \lambda\rho e_{t-1}^* = \varepsilon_t - v_{t-1} + \lambda\rho v_{t-2}. \quad (19)$$

The right-hand side of Eq. (19) is the sum of an MA(1) process and white noise, and therefore also an MA(1) process (see, e.g., [Hamilton, 1994](#), pp. 102–105). Denote the resulting process as $\xi_t + \theta\xi_{t-1}$ for some parameters θ and σ_ξ^2 to be determined. For the representation to be valid, the autocovariances must match, namely

$$\begin{aligned} \sigma_\varepsilon^2 + [1 + (\lambda\rho)^2]\sigma_v^2 &= (1 + \theta^2)\sigma_\xi^2 \\ -\lambda\rho\sigma_v^2 &= \theta\sigma_\xi^2 \end{aligned}$$

Substituting out σ_ξ and rearranging leads to a quadratic equation in θ :

$$(-\lambda\rho)\theta^2 - \theta \left\{ \frac{\sigma_\varepsilon^2}{\sigma_v^2} + [1 + (\lambda\rho)^2] \right\} - \lambda\rho = 0.$$

The equation has two real solutions. Picking the solution associated with the invertible representation (i.e., with $|\theta| < 1$) yields

$$\theta = \frac{\left\{ \frac{\sigma_\varepsilon^2}{\sigma_v^2} + [1 + (\lambda\rho)^2] \right\} - \sqrt{\left\{ \frac{\sigma_\varepsilon^2}{\sigma_v^2} + [1 + (\lambda\rho)^2] \right\}^2 - 4(\lambda\rho)^2}}{-2\lambda\rho}. \quad (20)$$

All in all, the observed forecast errors follow an ARMA(1, 1) process with

$$e_t^*(1 - \lambda\rho L) = (1 + \theta L)\xi_t,$$

where L is the lag operator. As a result, the Wold representation of e_t^* is

$$e_t^* = \xi_t + (\lambda\rho + \theta) \sum_{\ell=1}^{+\infty} (\lambda\rho)^{\ell-1} \xi_{t-\ell}.$$

Therefore, measurement error leads to an attenuation bias. The attenuation bias can be substantial if measurement error is large (i.e., signal-to-noise ratio, $\sigma_\varepsilon/\sigma_v$, is small). For example, suppose that $\lambda = 0.50$, $\rho = 0.75$, $\sigma_\varepsilon = 0.25$, and $\sigma_v = 0.15$. Then, the true bias coefficient b_1 is equal to $(-\lambda\rho) = -0.375$. In contrast, the bias coefficient in the process with measurement error is equal to

$$-(\lambda\rho + \theta) \approx -(0.375 - 0.097) = -0.278.$$

The attenuation bias is roughly 26% in relative terms. This simple calculation highlights the importance of having datasets in which many forecasters forecast the same variable. By averaging across multiple forecasters, σ_v can be reduced, thereby diminishing the attenuation bias.

Now consider the case of diagnostic expectations. In that case, the true forecast errors follow an MA(1) process with $e_t = \varepsilon_t - \theta\rho\varepsilon_{t-1}$, implying that the observed forecast errors are given by $e_t^* = \varepsilon_t - \theta\rho\varepsilon_{t-1} - v_{t-1}$. The right-hand side again follows an MA(1) process but with different parameters. Write $e_t^* = \xi_t + \psi\xi_{t-1}$ for some parameters ψ and σ_ξ^2 . Similar calculations to those performed earlier show that

$$\psi = \frac{\left\{ \frac{\sigma_v^2}{\sigma_\varepsilon^2} + [1 + (\theta\rho)^2] \right\} - \sqrt{\left\{ \frac{\sigma_v^2}{\sigma_\varepsilon^2} + [1 + (\theta\rho)^2] \right\}^2 - 4(\theta\rho)^2}}{-2\theta\rho}.$$

To gauge the size of the attenuation bias, suppose that $\theta = 0.50$, $\rho = 0.75$, $\sigma_\varepsilon = 0.25$, and $\sigma_v = 0.15$. With these parameters, the true bias coefficient is equal to $b_1 = \theta\rho = 0.375$. However, the bias coefficient from the process with measurement error (i.e., $-\psi$) is equal to approximately 0.268. In relative terms, the attenuation bias is roughly 29%.

B.2 Calibrating the Level of Measurement Error

We now study how large measurement error is likely to be in the context of inflation expectations. As a side benefit, the section characterizes the effects of measurement error on the procedure proposed by [Coibion and Gorodnichenko \(2015\)](#).

[Coibion and Gorodnichenko \(2012\)](#) use an empirical strategy that directly estimates the underlying ε_t shocks. They find (p. 154, an average across all their specifications) that $\hat{\lambda} = 0.82$. In subsequent work, [Coibion and Gorodnichenko \(2015\)](#) use a different estimation strategy to measure λ and find that $\hat{\lambda} = 0.54$ (p. 2653). The second strategy is potentially subject to attenuation bias stemming from measurement error (see their Footnote 7 and Online Appendix A).

Suppose that—as an upper bound calculation—we attribute all of the difference between the two estimates to measurement error in the second approach. [Coibion and Gorodnichenko \(2015\)](#) obtain λ by estimating the following regression by least squares:

$$\underbrace{x_t - \mathbb{F}_{t-1}^*[x_t]}_{\equiv e_t^*} = \alpha + \beta \{ \mathbb{F}_{t-1}^*[x_t] - \mathbb{F}_{t-2}^*[x_t] \} + u_t.$$

The stars indicate that the variables are measured with error:

$$\begin{aligned} \mathbb{F}_{t-1}^*[x_t] &= \mathbb{F}_{t-1}[x_t] + v_{t-1} \\ \mathbb{F}_{t-2}^*[x_t] &= \mathbb{F}_{t-2}[x_t] + v_{t-2} \end{aligned}$$

where v_t is a white noise measurement error with variance σ_v^2 . As a result, $e_t^* = e_t - v_{t-1}$ (with e_t denoting the one-step ahead forecast error).

For the sticky information model, k -step ahead forecasts are given by

$$\mathbb{F}_t[x_{t+k}] = (1 - \lambda) \sum_{\ell=0}^{+\infty} \lambda^\ell \mathbb{E}_{t-\ell}[x_{t+k}] = \rho^k \sum_{\ell=0}^{+\infty} \rho^\ell (1 - \lambda^{\ell+1}) \varepsilon_{t-\ell}, \quad (21)$$

implying that $\mathbb{F}_t[x_{t+2}] = \rho \mathbb{F}_t[x_{t+1}]$. Forecast revisions for x_{t+1} are hence

$$\mathbb{F}_t[x_{t+1}] - \mathbb{F}_{t-1}[x_{t+1}] = \rho(1 - \lambda) \sum_{\ell=0}^{+\infty} (\rho\lambda)^\ell \varepsilon_{t-\ell}.$$

Since the forecast errors are equal to $e_t = \sum_{\ell=0}^{+\infty} (\rho\lambda)^\ell \varepsilon_{t-\ell}$, we can write

$$e_t = \varepsilon_t + \frac{\lambda}{1 - \lambda} \left[\rho(1 - \lambda) \sum_{\ell=0}^{+\infty} (\rho\lambda)^\ell \varepsilon_{t-1-\ell} \right] = \varepsilon_t + \frac{\lambda}{1 - \lambda} \{ \mathbb{F}_{t-1}[x_t] - \mathbb{F}_{t-2}[x_t] \}.$$

Combining these results we have that

$$\text{Var}\{\mathbb{F}_{t-1}[x_t] - \mathbb{F}_{t-2}[x_t]\} = \frac{\rho^2(1-\lambda)^2\sigma_\varepsilon^2}{1-\lambda^2\rho^2},$$

and therefore

$$\text{plim } \hat{\beta} = \frac{\lambda(1-\lambda)\rho^2 - \frac{\sigma_v^2}{\sigma_\varepsilon^2}(1-\lambda^2\rho^2)}{(1-\lambda)^2\rho^2 + 2\frac{\sigma_v^2}{\sigma_\varepsilon^2}(1-\lambda^2\rho^2)}.$$

Note that there are *two* reasons why the empirically estimated $\hat{\beta}$ would understate the true level of stickiness. First, there is the usual attenuation bias, as seen in the denominator. However, measurement error also introduces a mechanical negative correlation between forecast errors and past forecast revisions, as captured by the second term in the numerator. As a result, even a small amount of measurement error can lead to substantial bias. If measurement error is severe, the estimated $\hat{\beta}$ may even be negative. Indeed, if $\sigma_v \rightarrow +\infty$, $\text{plim } \hat{\beta} \rightarrow -1/2$. We emphasize that measurement error can also be given a more structural economic interpretation. For instance, measurement error is isomorphic to some models of overconfidence (Section A.6).

We can now ask at what level of measurement error the probability limit above would coincide with the empirically measured $\hat{\beta}$. Coibion and Gorodnichenko (2015, p. 2653) estimate $\hat{\beta} = 1.19$. Take $\rho = 0.85$, $\sigma_\varepsilon^2 = 1.005$, and $\lambda = 0.82$ (using the estimates in Coibion and Gorodnichenko, 2012). Solving for σ_v^2 , we find that $\sigma_v \approx 0.21$. In other words, for the estimates in Coibion and Gorodnichenko (2012) and Coibion and Gorodnichenko (2015) to be consistent, the magnitude of measurement error needs to be roughly a fifth of the magnitude of the true shocks.

Appendix C Data Appendix

We download data for the individual responses from the website of the Federal Reserve Bank of Philadelphia ([link](#)). The downloaded file contains forecasts of GDP deflator inflation for the past quarter ($PGDP1$), current quarter ($PGDP2$), and the next four quarters ($PGDP3$ up to $PGDP6$), see [Federal Reserve Bank of Philadelphia \(2017, pp. 20–22\)](#).

To construct consensus inflation forecasts, we first calculate the median forecast of $PGDP2$ and $PGDP3$ in each quarter. Then, we calculate annualized quarter-on-quarter inflation forecasts as

$$100 \left[\left(\frac{PGDP3}{PGDP2} \right)^4 - 1 \right].$$

The approach follows the standard practice in the Survey of Professional Forecasters. To calculate individual inflation forecasts, we directly use the equation above.

For realizations, we use the Real-Time Data Set for Macroeconomists which is also provided by the Philadelphia Fed ([link](#)). We use the first-release data for “*Price Index for GNP/GDP (P)*.” In 1995Q4, the first-release data for inflation is not available. In this period, we use the second-release data.

To match forecasts and actuals, we align the forecasts to the date for which they were made. For example, the one-quarter ahead forecast made in the 1970Q1 survey is matched with the actual inflation reported for 1970Q2.

Appendix D Additional Results

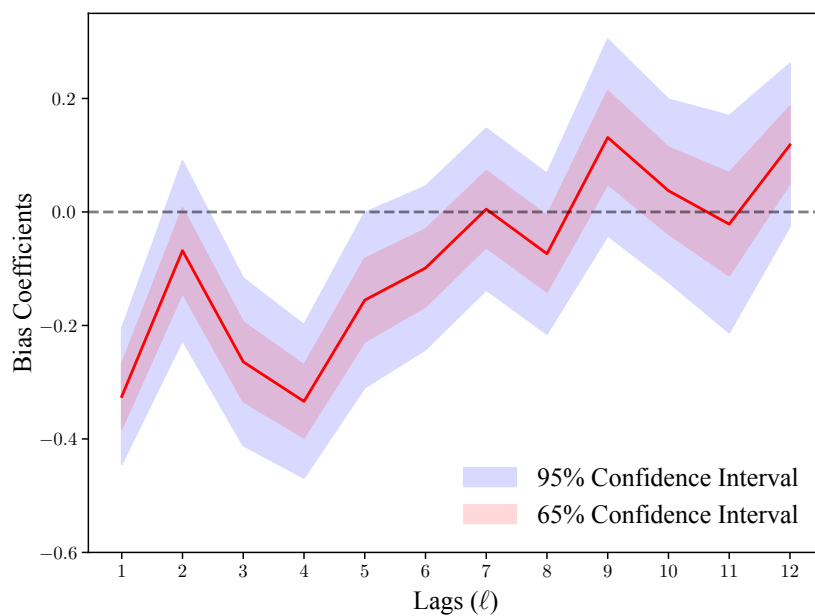


Figure 7: Bias coefficients for one-quarter ahead inflation forecasts: maximum likelihood estimates. The estimation uses consensus (median) forecasts. The impulse response function of the forecast errors is obtained by estimating

$$x_t - \mathbb{F}_{t-1}[x_t] = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_{12} \varepsilon_{t-12}, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

by maximum likelihood. The bias coefficients are then given by $\hat{b}_\ell = -\hat{\theta}_\ell$.

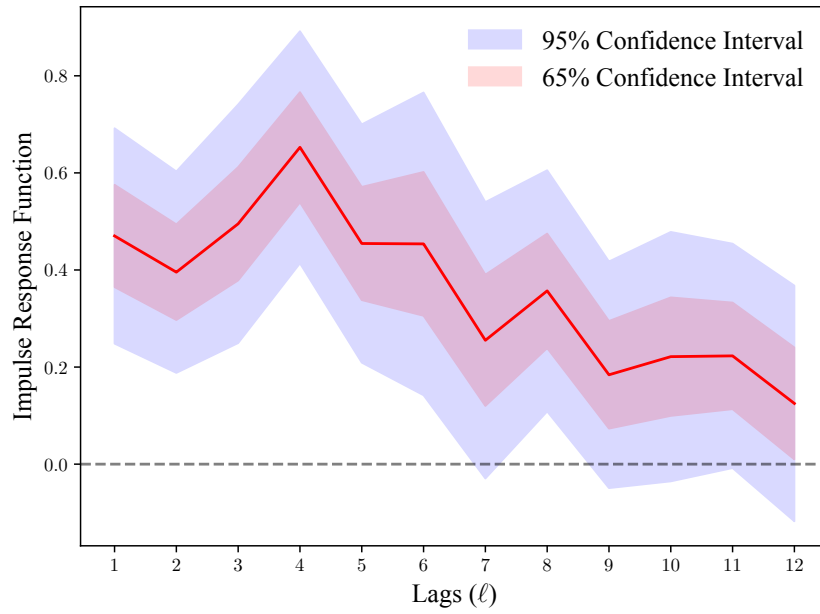


Figure 8: Impulse response function: quarterly GDP deflator inflation. The estimation uses local projections; Newey-West standard errors with four lags are used to calculate the confidence intervals.