

Comment

Jessica A. Wachter*

June 30, 2020

I Introduction

The authors build a quantitative framework, broad enough to nest different types of ideas proposed in behavioral macroeconomics, to help make sense of the survey evidence on expectations. The goal is not so much to understand survey evidence for its own sake, though that might be interesting. Nor is it to decide between a class of competing models, all of which work fairly well. Rather, the goal is to inform future directions for research, in a field which, at the moment seems a bit lost in a “wilderness.” In the end, using their approach, the authors are able to say that the weight of the evidence favors some types of models rather than others.

The focus on survey evidence comes at a time of uncertainty about future directions in macroeconomics. Many, including some of the authors in prior work, have noted serious failings with benchmark models. These failings have a solution, perhaps, within a different class of models, ones that do not impose rational expectations. But rational expectations is a powerful disciplining device and this broader class seems impossibly large.

*The Wharton School, University of Pennsylvania

The literature has, for several years, recognized the potential of survey evidence to itself be a disciplining device, filling some of the role that rational expectations used to, and does still, play. Barberis et al. (2015), for example, argue for using survey evidence to decide between competing models in financial economics. In macroeconomics, Coibion and Gorodnichenko (2015) argue that one need not have a model to interpret survey evidence: a positive sign in regressions of forecast errors on forecast revisions is model-free evidence of under-reaction in macroeconomics, relative to a rational expectations benchmark.

Coibion and Gorodnichenko (2015) indeed find evidence for such under-reaction. Their preferred interpretation is one of dispersed private information. The authors discuss a substantial literature in macroeconomic theory that proposes other types of under-reaction operating at the level of the individual. Indeed, dispersed private information is a form of individual under-reaction because it requires some individuals to persist in having different views than others in what would seem in principle to be the same data.

In fact, under-reaction might be the first model that would come to mind if one moved away from the rational expectations benchmark. Suppose that some process interfered with agents' encoding of information (one could call it limited attention), or that agents forgot information with some probability (formalized by Mullainathan (2002)). In either type of model, beliefs would display persistent changes due to the fact that it takes several "tries" for new information to enter the agent's database.

II Summary of the evidence

To understand the results in this paper and the related literature, it is first helpful to review the terminology. Agents forecast a macroeconomic series x_t . Let \mathbb{E}_t denote the forecast as of time t – note that, like the authors, I assume by use of this notation that

the forecast is an expectation (and not, say, a median) – so that $\mathbb{E}_t[x_{t+k}] - \mathbb{E}_{t-1}[x_{t+k}]$ is the revision in expectation as of time t , whereas $x_{t+k} - \mathbb{E}_t[x_{t+k}]$ is the forecast error. The idea behind the Coibion and Gorodnichenko (2015) regressions is to attempt to forecast the error with the revision. Thus the revision in the forecast is used to forecast the forecast error. This is a mouthful, so I will simply refer to the right hand side variable as the revision, and the left hand side variable as the error.

Assuming \mathbb{E}_t is a true expectation, namely it obeys the laws of conditional probability, then the error should be unforecastable by any variable available to the agent at time t . Presumably, this should include the agent’s own revision. After all, if something can forecast the error, it follows that it can be used to improve the expectation, and should be part of the revision.

What if, however, there is information at time t not captured in the forecast? Bordalo et al. (2018) argue for running the regression on individual-level data. They point out, as do Coibion and Gorodnichenko (2015), that forecastability in of errors from median forecasts is compatible with individual rationality (at least in a narrow sense). If some individuals lack information that goes into the median forecast, what looks like predictability will appear to be a surprise to these investors, who are late in the game. This situation is called “dispersed private information,” and, while compatible with Bayesian updating, it requires artificially restricting the data, or manipulation of the priors.

Indeed, it is hard to argue that regressions with \mathbb{E}_t representing an individual forecast does not represent rationality, at least if we consider rationality as equivalent to Bayesian updating (and if we assume that the forecast is a mean, and not, say a median). Bordalo et al. (2018) show that the individual-level forecasts are a mix of over and under-reaction, but with over-reaction dominating. The importance of this result is best understood in light of the goals in the first paragraph. Bordalo et al. (2018) argue that over-reaction is the trace we see of the representative heuristic at

work. The representative heuristic, though it has no agreed-upon cognitive foundation, is a candidate for a unifying principle in behavioral macroeconomics.

The authors rerun the regressions from these prior papers (Tables 1 and 2). Confirming results from the prior literature, they find that the coefficient in median-level forecasts (K_{CG}) for unemployment and inflation is positive – consistent with under-reaction.¹ They find a similar result for the coefficient on individual-level forecasts (K_{BGMS}) for unemployment, but not for inflation. For inflation, results for K_{BGMS} are inconclusive for the full sample and strongly negative – indicating over-reaction – for a subsample beginning in 1984.

III Summary of novel findings

One message of the paper is that the survey data are indeed informative, but not in the way we might first think. The paper offers a summary of evidence and additional tests, then a model that reconciles seemingly contradictory findings in the survey evidence, taking the DGP as given.

Finally, the authors go a step further and derive what would the DGP be in a New Keynesian model in which agents' beliefs followed the specified form. The authors then have a truly general equilibrium model of survey evidence, perhaps the first of such a kind. While this is impressive, I can't help but wonder if this second part of the paper may have less of an impact than the first. After all, the first part potentially speaks to anyone with an interest in how beliefs are formed, whereas the second part does require subscribing to the New Keynesian framework. Yes, the first part does formally require specification of a DGP; yet here the authors make a fairly natural and parsimonious choice. Moreover, their approach could be applied to other DGPs.

¹A positive coefficient indicates under-reaction because the agents do not update enough, given whatever information made them revise their estimates.

A New evidence

The authors augment the survey evidence with evidence from a VAR using shocks to unemployment and inflation. Consider :

$$x_t = \alpha + \sum_{i=1}^I \gamma_i x_{t-i}^{IV} + \beta_0 \epsilon_t + u_t$$

where x_t could be the underlying series or its median forecast, and ϵ_t is a shock, and x_{t-i}^{IV} are instrumented regressors, so γ_i correctly measures the response of x_t to x_{t-i} .

Figure 3 in the paper shows the results: the forecasts clearly lie below the outcome at the start (under-reaction), and then above the outcome at the end. It looks like under-reaction, and then over-shooting. This is true for both unemployment and inflation. The degree of over-shooting (the fact that expectations start out below the realized and then end up above) is captured in a coefficient K_{IRF} , which is greater than one if over-shooting occurs. The goal of the model will be to jointly match K_{CG} , K_{BGMS} and the novel VAR evidence (K_{IRF}).

Already, however, a tension appears. Though estimated on the same data as Coibion and Gorodnichenko (2015), the authors have found evidence for a type of over-reaction in expectations, namely over-shooting. Coibion and Gorodnichenko (2015) show that median forecasts under-react. Figure 3 shows that median forecasts over-shoot. How can both be true?

A careful look at Figure 3 does reveal pervasive under-reaction. At the start, the forecast lies below the outcome. Both the forecast and the outcome are rising. But the forecast does not rise fast enough, and so, in the short-term errors are positive. Then the forecast crosses the outcome. Now the outcome is below the forecast. Both the outcome and the forecast are falling. But the forecast does not fall fast enough. Thus the forecast does under-react in the short-term. One contribution of the paper – is the finding that there are situations in which a negative coefficient in Coibion and

Gorodnichenko (2015) regressions is consistent with long-run over-reaction.

B The model confronts the evidence

The model can easily be described in a few short equations. assume the following data generating process for x_t :

$$(1 - \rho L)x_t = \epsilon_t \sim N(0, 1).$$

Agent i observes signal $s_{i,t}$, equal to x_t plus normally distributed noise:

$$s_{i,t} = x_t + \frac{u_{i,t}}{\sqrt{\tau}}.$$

This is dispersed private information: agents observe their own signals and cannot share them with others. The agent will form beliefs using a standard Kalman filter, except that the agent's beliefs are mis-specified. In particular, the agent believes:

$$(1 - \hat{\rho}L)x_t = \epsilon_t$$

and understands that the signal is noisy, but mis-judges the noise of signal:

$$s_{i,t} = x_t + \frac{u_{i,t}}{\sqrt{\hat{\tau}}}.$$

Under this model, the coefficient for individual forecasts, up to a constant of proportionality, equals:

$$K_{BGMS} \propto -\kappa_1(\tau^{-1} - \hat{\tau}^{-1}) + \kappa_2(\rho - \hat{\rho}).$$

The evidence is that this is < 0 for inflation and > 0 for unemployment. Given the coefficient for individual forecasts, the coefficient for the aggregate forecast is proportional

to

$$K_{CG} \propto \kappa_1 \tau^{-1} + V_{\text{ind}} K_{BGMS}$$

The evidence is that that is > 0 . Finally, the measure of over-shooting equals

$$K_{IRF} = \frac{\log(\hat{\rho} - \rho) - \log(\hat{\rho} - \hat{\lambda})}{\log \hat{\lambda} - \log \rho}$$

The evidence says that this is > 1 .

In light of the model, how can we interpret this evidence? First, $K_{CG} > 0$ indicates dispersed private information. Unlike other types of under-reaction that the literature has proposed, dispersed private information is compatible with beliefs over-shooting. Second, $K_{IRF} > 0$ implies over-extrapolation, not over-confidence. While $K_{BGMS} < 0$ by itself might imply over-confidence, over-confidence operates on impact, and one needs some mechanism that operates on long-run expectations.

Thus dispersed private information becomes the primary mechanism for $K_{CG} > 0$, and over-extrapolation, the primary mechanism for $K_{BGMS} < 0$. Because over-extrapolation is doing the work of fitting both K_{BGMS} and K_{IRF} (over-shooting) “confidence” (over or under-) then becomes the “plug” for which one fit K_{BGMS} if necessary. If one wonders whether this extra degree of freedom – confidence – means that we cannot identify K_{CG} , one need not worry, because confidence and dispersed private information can then be separately identified from the differences in K_{CG} and K_{BGMS} .

C Comments on the model

A general comment is that these agents are fairly rational in that their beliefs are governed by Bayes rule. Putting them into an economy with asset prices, implies that asset prices would obey no-arbitrage. They do not, however, display rational expectations in the sense of “communism of models” (Hall and Sargent, 2018). The

true data generating process (DGP) does not equal the subjective DGP, and moreover, agents cannot learn the true DGP because they have, in a sense, dogmatic priors over $\hat{\tau}$ and $\hat{\rho}$. I give several specific comments below.

What is the correct null hypothesis?

The equations for K_{CG} , K_{BGMS} , and K_{IRF} have the required interpretation only if the agents are running the same Kalman filter as the authors. On one level, this is a restatement of the fact that they are Bayesian. But they are required to solve a complicated filtering problem in which mistakes are possible. For example, the agents may make sub-optimal use of the previous signal (or may fail to use this signal entirely). If the agent simply bases today's forecast on today's information, which does not, intuitively, seem that far from optimality, then $K > 0$.

More generally, mistakes would seem to point to $K < 0$. All that is needed is that large deviations from a fundamental are more likely to be wrong. This is because large deviations are likely to be on the same side of both the truth and previous forecast. If the agent suddenly forecasts a high value for unemployment, say, this will be coded as an upward revision. When unemployment is turns out to be low relative to the forecast, the error is negative. Potential sources of error are numerous: they could be cognitive (the agent has the wrong model), perhaps a computer bug, perhaps an error in the agent's own data, perhaps an error in reporting the forecast, or in entering the forecast into the database.² As does the field at large, the authors accept the existence of such errors when following the common practice of Winsorizing the data. But at what point is something an "outlier" that must be removed, versus an error indicating, say, over-confidence or over-extrapolation? The answer is not obvious.

²One might respond: but this error is exactly what we are trying to pick up! But in fact this is not the case. The ultimate goal is to model economic behavior, because this is what matters for macro-economic aggregates. Evidence suggests that individuals do not take their own beliefs seriously (Giglio et al., 2019).

Confidence: over or under?

The situation is further complicated by the fact that the cited evidence for over-or under-confidence does not support the usual interpretation of these terms in economics. Moore and Healy (2008) show that, in difficult tasks, subjects over-estimate their own performance, while at the same time believing that they are worse than others at the tasks. For easy tasks, subjects under-estimate their own performance, but believe they are better than others. One interpretation is that agents are sometimes over- and sometimes under-confident.

But a more parsimonious interpretation is that agents shade their own judgements toward the mean. When they perform badly, they interpret the feeling of “performing badly” as in part that the task was difficult for them (they are worse than others), but also in part that they performed better than they thought (namely, they understand that their sense of performing badly is subjective to error). When they perform well, they interpret that subjective feeling as the task being easy (they are better than others), but also, that their own judgement may be in error. Agents are doing what we say they should under a model where they receive a noisy signal – weighting that signal toward the mean.

One might object: why can't the agent's correctly sense their own performance and also simply accept that the tasks are hard or easy for everyone? But in the Moore and Healy (2008) experiment, the agents do not know the “true” difficulty of the task. In world we are trying to model, agents may not know the true dispersion of private information, which would correspond to the true task difficulty. But there is no particular reason to believe that they would not try to learn in, or even eliminate said private information. From a cognitive point of view, under or over-confidence as a parameter is suspect.

What is over-extrapolation?

In the model, over-extrapolation is $\hat{\rho} > \rho$. The literature on behavioral economics has proposed two very different foundations for this behavior. The first is simply: extrapolation. Extrapolation could arise from recency effects (agents are more likely to remember recent events). This is well-established in laboratory experiments. It is also well-established in the data when it comes to understanding stock returns, a setting where it is distinctly suboptimal (Greenwood and Shleifer, 2014). In any case, it is clear that the recent past matters in decision-making beyond what it should in a Bayesian model.

A second proposed mechanism, and the focus of Bordalo et al. (2018), is the representativeness heuristic. The idea behind the representativeness heuristic comes from a literature on judgement and decision-making in psychology, and in particular from Tversky and Kahneman (1973). In their work, agents, when asked to make a probabilistic judgement (e.g. how likely is it that an individual, certain information, works in a certain occupation), will give a high value if the individual “matches” the category – represents essential features of the category – and ignore the base rate of the occupation. Bordalo et al. (2018) apply representativeness to the problem of detecting an underlying process given a signal. In their work, the signal becomes representative of a latent state, and agents over-estimate its value in saying something about that state. In the present paper, the representative heuristic is over-confidence: the agent updates too much on impact. The paper firmly rejects this viewpoint of “over-reaction on impact.”

Evidence of some type of over-extrapolation is present in asset prices.³ Figure 1 is

³One can of course think of asset prices as a type of survey data because they reflect the (risk-neutral) forecasts of investors. While much attention has been paid to the difference between risk-neutral and physical forecasts, it is not obvious that the actual difference warrants the attention. Moreover, conceivably survey evidence also suffers from this problem as forecasters may be reporting what agents most scare about.

from recent work (Guo and Wachter, 2019). This figure shows the time series of prices and dividends. Updating work of Shiller (1981) and Campbell and Shiller (1988) to the present, this figure shows at least graphical evidence that when prices and dividends diverge, it is prices that fall to meet dividends and not the other way around.

Interpreting the inflation series

Figure 2 shows the median forecast from the SPF as well as realized inflation from the data. Recall that the initial interpretation of expectations was under-reaction (Coibion and Gorodnichenko, 2015). Then, it became over-reaction (Bordalo et al., 2018), via the representativeness heuristic. The current paper interprets it as over-shooting (agents over-estimate the persistence). Looking at the figure reveals the potential for confusion.

The figure reveals that the forecast clearly missed the high inflation in the 1970s, nor did it catch up sufficiently quickly. Then it missed the speed of the downturn, and again, it did not catch up sufficiently quickly. Remarkably, there was a period over a decade long: early-1980s until the late 1990s, in which for almost every month, the median forecast lay above the actual realized inflation.

One could interpret this as over-shooting: agents were slow to react to the shock, and then, the face of much evidence, continued to believe that inflation would be high. Figure 2 suggests, however, that the VAR is mis-specified.⁴ The problem plaguing much research in expectations comes in full force. To interpret behavior as irrational, we need a benchmark model. If the benchmark model is mis-specified however, our interpretation is subject to error and beliefs may be more rational than the economists' specifications. This way of thinking can lead to an infinite regress because (a) one

⁴An interesting area for further work is the interpretation of the impulse response functions in light of such mis-specification. Much research has shown a bias in autocorrelations stemming from uncertainty in the mean. It may be that the results from Figure 3 are sensitive to whether one imposes that the forecast and the actual data have the same mean. Moreover, if either is non-stationary, then the figure is uninterpretable. A Monte Carlo study under various plausible DGPs would answer that question.

always needs a model to interpret expectations and (b) *any* model is wrong. There is no solution to this problem except perhaps to remember that terms like rationality are ultimately technical (Bayes law, for example, or “communism of beliefs”), and should not carry their normal English-language associations.

This state of confusion be unimportant if one could just lay aside the forecasts of inflation. However, expectation formation of inflation has been a central agenda in macroeconomics since the very beginning of the field. Accounts of Federal Reserve decision-making suggests that it operates with implicit knowledge of expectations formation (Tarullo, 2017). Closing the gap between this implicit knowledge and formal knowledge through science seems important.

IV Conclusion: what should a model eventually look like?

This paper stresses dispersed private information, under-confidence, and over-extrapolation. But the underlying parameters, τ , $\hat{\tau}$, $\hat{\rho}$, must be estimated series-by-series. Why might these differ in different series? Do these parameters, for instance, form the correct explanation for, say, the the failure of forward guidance?

Here is an alternative view of what a model could look like: First, a growing body of literature suggests experience effects persist for far longer than they should.⁵ The series itself suggests that forecasters were permanently influenced by the high inflation in the 1970s. A benefit of a model of experience effects is that it endogenizes “private information.” It is not so much that information is private, but rather than agents filter that information differently through the lens of their experiences.

Second, results in this paper suggest a slow updating to new information, with eventual over-shooting. In recent work, (Wachter and Kahana, 2020), Michael Kahana

⁵On inflation, see, for example, Malmendier and Nagel (2016).

argue that experience effects and slow updating to new information, with over-shooting, can be captured by incorporating principles of association, together with encoding and retrieval of a persistent contextual state. Laboratory tasks on free recall show that agents possess a persistent mental context. Context and an endogenous set of associations influence what comes to mind. Context then responds to features of the environment based on these associations (Howard and Kahana, 2002). Incorporating these ideas into a model of financial decision-making, we show that agents can initially under-react to novel features of the environment. If the novel features display some permanence, beliefs will update too much (over-shooting) in that agents will forget the previous features. Perhaps most importantly, though, the model incorporates long-run experience effects through prior associations. These associations continue to influence how the agent sees the data.

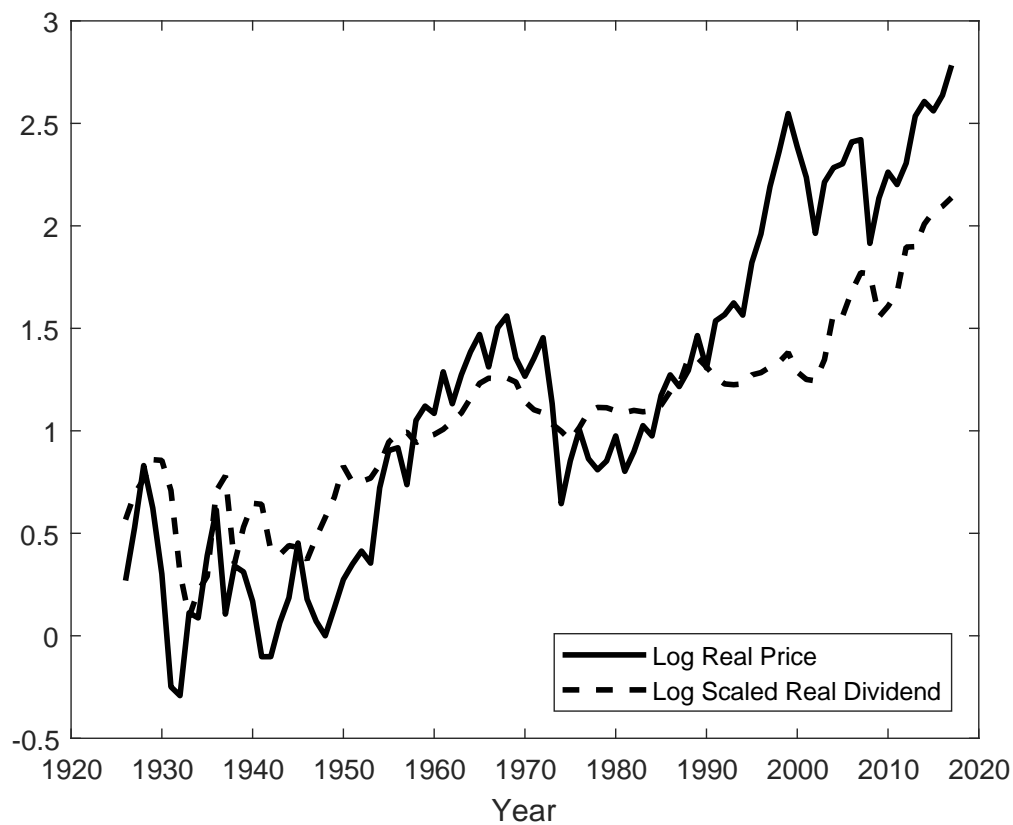
This paper has indeed helped to show the way with new results and a new theoretical framework. If one loses something of the cognitive foundations, one certainly gains in terms of tractability and generality.

References

- Barberis, N., Greenwood, R., Jin, L., and Shleifer, A. (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics*, 115(1):1–24.
- Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2018). Over-reaction in macroeconomic expectations. NBER Working Paper No. 24932.
- Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.
- Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–2678.
- Giglio, S., Maggiori, M., Stroebel, J., and Utkus, S. (2019). Five facts about beliefs and portfolios. NBER Working Paper No. 25744.
- Greenwood, R. and Shleifer, A. (2014). Expectations of returns and expected returns. *Review of Financial Studies*, 27(3):714–746.
- Guo, H. and Wachter, J. (2019). ‘superstitious’ investors. NBER Working Paper No. 25603.
- Hall, R. E. and Sargent, T. J. (2018). Short-run and long-run effects of Milton Friedman’s presidential address. *Journal of Economic Perspectives*, 32(1):121–134.
- Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3):269–299.
- Malmendier, U. and Nagel, S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1):53–87.

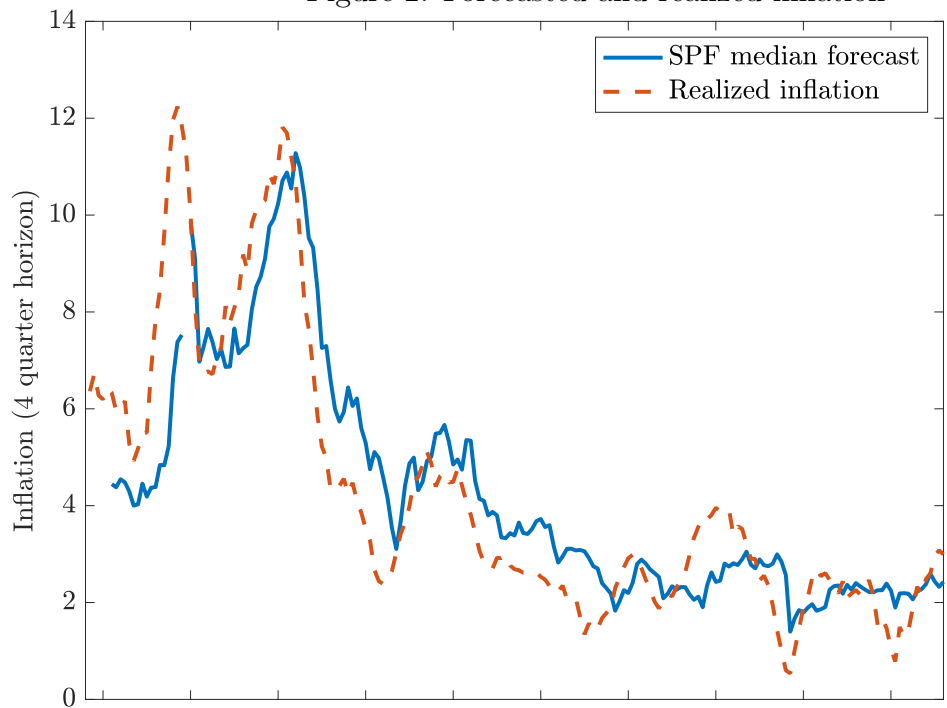
- Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2):502–517.
- Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*, 117(3):735–774.
- Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71:421–436.
- Tarullo, D. K. (2017). Monetary policy without a working theory of inflation. Hutchins Center Working Paper #33.
- Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232.
- Wachter, J. A. and Kahana, M. J. (2020). A retrieved-context theory of financial decisions. NBER Working Paper No. 26200.

Figure 1: Log Real Dividend and Log Real Price Level of the Aggregate Market, Full Sample



This figure plots the annual frequency log price level and log dividend of the US stock market in the post-1926 era. The log real dividend is multiplied by 27.43, the mean P/D ratio post-1926. The dividend and the price are adjusted for inflation.

Figure 2: Forecasted and realized inflation



The figure shows the median inflation forecast from the survey of professional forecasters, as well as realized inflation.