ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Asymmetry, Complementarities, and State Dependence in Federal Reserve Forecasts

Authors	Julieta Caunedo, Riccardo DiCecio, Ivana Komunjer, and Michael T. Owyang
Working Paper Number	2013-012C
Revision Date	January 2020
Citable Link	https://doi.org/10.20955/wp.2013.012
Suggested Citation	Caunedo, J., DiCecio, R., Komunjer, I., Owyang, M.T., 2020; Asymmetry, Complementarities, and State Dependence in Federal Reserve Forecasts, Federal Reserve Bank of St. Louis Working Paper 2013-012. URL https://doi.org/10.20955/wp.2013.012

Published In	Journal of Money, Credit, and Banking
Publisher Link	https://doi.org/10.1111/jmcb.12590

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Asymmetry, Complementarities, and Federal Reserve Forecasts*

Julieta Caunedo† Riccardo DiCecio‡ Ivana Komunjer§ Michael T. Owyang¶

December 29, 2017

Abstract

Forecasts are a central component of policymaking; the Federal Reserve's forecasts are published in a document called the Greenbook. Previous studies of the Greenbook's inflation forecasts have found them to be rationalizable but asymmetric if considering particular subperiods—e.g., before and after the Volcker appointment. In these papers, forecasts are analyzed in isolation, assuming policymakers value them independently. We analyze the Greenbook forecasts in a framework in which the forecast errors are allowed to interact. We find that allowing the losses to interact makes the unemployment forecasts virtually symmetric, the output forecasts symmetric prior to the Volcker appointment, and the inflation forecasts symmetric after the onset of the Great Moderation.

Keywords: forecast rationality, loss function, Taylor rule, Greenbook forecasts, break tests.
JEL codes: C53; E47; E52.

*The paper was previously circulated as "Federal Reserve Forecasts: Asymmetry and State-Dependence". The authors thank Mengshi Lu, Mike MCCracken, and Chris Otrok for comments and suggestions. Diana A. Cooke, Hannah G. Shell, and Kate Vermann provided research assistance. Views expressed here are the authors' alone and do not necessarily reflect the opinions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†Department of Economics, Cornell University. Email: julieta.caunedo@cornell.edu.
‡Research Division, Federal Reserve Bank of St. Louis. Email: dicecio@stls.frb.org.
§Department of Economics, Georgetown University. Email: ik334@georgetown.edu.
¶Research Division, Federal Reserve Bank of St. Louis. Email: owyang@stls.frb.org.

1 Introduction

The forecasts that policymakes rely on to set policy are created as much on judgment and belief than on economic or econometric models. In a speech made on July 10, 2007, then-Federal Reserve Chairman Ben Bernanke noted[1]:

> The Board staff employs a variety of formal models, both structural and purely statistical, in its forecasting efforts. However, the forecasts of inflation (and of other key macroeconomic variables) that are provided to the Federal Open Market Committee (FOMC) are developed through an eclectic process that combines model-based projections, anecdotal and other "extra-model" information, and professional judgment.

These sentiments were echoed by Tarullo (2017), who suggested that "many of the concepts invoked by monetary policy, (...), are more directional indicators in assessing the economy than guides to individual policy decisions." Beliefs of the FOMC members about structural relationships among macro variables are key to the decision making process of the policy authority. Indeed, Sargent et al. (2006) formalizes how time-varying beliefs of the monetary authority can affect inflation outcomes in the economy, and show how the Fed's belief in a Phillips curve relationship could account for Volcker's disinflation.

Given the importance of beliefs in economic relationship for forecasting—and the possibility that these beliefs can change over time, one might wonder whether or how the Federal Reserve weighs these relationships and whether these beliefs are fluid. Even long standing macroeconomic relationships such as Okun's law and the Phillips curve have come into question by members of the FOMC and the Board staff. For example, On March 26, 2012, then-Federal Reserve Chairman Ben Bernanke cited what he dubbed the "Okun's law puzzle", a divergence from the historical statistical relationship between output and unemployment.[2] Brayton et al. (1999) argue that the standard view of the Phillips curve, which documents the relationship between unemployment and inflation, has changed over time.

The forecasts that the Fed produces to conduct monetary policy are contained in the Greenbook (or, more recently, the Tealbook). Made publicly available at a 5-year lag, Greenbook forecasts

[1] The full text of the speech is available here: https://www.federalreserve.gov/newsevents/speech/bernanke20070710a.htm.

[2] The full text of the speech and Bernanke's proposed explanations for the Okun's law puzzle can be found here: https://www.federalreserve.gov/newsevents/speech/bernanke20120326a.htm.

have been used to study various aspects of Fed behavior, such as whether the Fed's forecasts are rationalizable or whether the Fed has an informational advantage over the private sector.[3] The Fed's forecasting behavior is important for our understanding of monetary policy design. For example, previous studies have argued that observed shifts in Fed's forecasting behavior relate to changes in monetary policy objectives (see Orphanides, 2002; Capistrán, 2008).

While we may not directly observe the forecasting process, one way to evaluate a policymaker's belief in certain economic relationships is to examine their forecasts and, in particular, the joint behavior of their forecasts. Across the literature, forecasting behavior is modeled as a loss minimization problem in which the loss for each variable of interest are often assumed to be symmetric, independent of other variables, and time invariant. Symmetry implies equal loss for forecast errors of the same size regardless of the sign. Independence implies that the forecast errors of each variable are weighed separately.

Previous studies have argued that forecaster's preferences may not necessarily be symmetric (see Elliott et al., 2005, 2008). For example, for the policymaker, overpredicting output growth may be a worse outcome than an underprediction. In this paper, we will argue that preferences may also not be independent—or separable (see Komunjer and Owyang, 2012). For example, simultaneously underpredicting inflation and overpredicting output growth could yield higher loss than overpredicting inflation and overpredicting output growth. Finally, the forecaster's preference may vary as the strength of the belief in the underlying economic relationships changes (see Capistrán, 2008).

In this paper, we study the Fed's forecasting behavior in an environment in which losses for forecast errors in inflation, unemployment rate, and output growth are asymmetric, interact with each other, and vary over time. Ellison and Sargent (2012) argue that Federal Open Market Committee (FOMC) members forecasts might be biased (and less efficient under quadratic loss) because robust policymaking under model uncertainty may require policymakers to make "worst case scenario" forecasts. If the risks are asymmetric (i.e., inflation above target and unemployment

[3] For example, Swanson (2004) finds that the Fed's forecast errors are, on average, biased and/or correlated with the information available at the time the forecasts are made. Standard tests—e.g., those employing Theil-Mincer-Zarnowitz regressions—reject that the Fed forecasts are rationalizable (e.g., Jansen and Kishan, 1996); others (e.g., Sims, 2002; Romer and Romer, 2000) have determined that Fed forecasts are not only rationalizable but encompass private sector forecasts. Other recent papers that examine the Greenbook forecasts are Wang and Lee (2012) and Rossi and Sekhposyan (2011).

above target are more costly than their counterparts), it makes sense that the forecasts could be biased.

We find evidence that the forecast preferences of the Fed are asymmetric but that these asymmetries can change over time. We find two breaks in the parameters of the forecast loss function. The first break is consistent with Capistrán (2008), who finds that the forecast behavior changes when Paul Volcker was appointed as Chairman of the Federal Reserve Board. The second break occurs at the end of the Volcker disinflation, coincident with the onset of the Great Moderation. We find that the losses associated with output growth forecast errors are symmetric prior to Volcker's appointment and that the losses associated with inflation forecast errors are symmetric during the Great Moderation. Thus, the post-Volcker loss asymmetries found for inflation forecasts in other studies are mostly accounted for by the Volcker disinflation period. We also show that nonseparability in the loss function is important: Losses associated with unemployment rate forecasts appear to be asymmetric on their own but prove to be virtually symmetric when we account for interactions between the unemployment rate and the forecast errors in either inflation or output.

Considering these nonseparabilities are important for determining how the policymaker weighs the strength of the underlying economic relationships. We find that the complementarities—computed as the second cross-derivative of the forecast loss with respect to a pair of variables—between the unemployment rate and output growth and between output growth and inflation have changed substantially across time. In particular, we find that, during the Great Moderation period, the above complementarities have disappeared or reversed sign. We interpret these changes as the forecaster attaching less weight to (and perhaps belief in) the relationships between the two variables.

We argue these results hint to an alternative interpretation of the historical analysis of monetary policy. Orphanides (2001, 2002, 2004) argues that starting with Chairman Volcker, the Fed came to terms with the difficulties in predicting real variables. Thus, the Fed became less active in trying to stabilize the perceived output gap. Our results are broadly consistent with this previous finding, but can be viewed from a slightly different angle based on the preferences of the forecaster. We conclude that the Fed forecasts changed from emphasizing two-sided accuracy in output growth to emphasizing two-sided accuracy in inflation.

The balance of the paper is outlined as follows. Section 2 reviews the multivariate non-separable asymmetric loss function and its properties. It also provides an illustrative example of symmetric/asymmetric and separable/non-separable loss. Section 3 describes the Greenbook forecast data, the realization data, and the instruments. This section also describes the estimation of the loss function parameters, the J-test used to evaluate forecast rationality, and the break test for the asymmetry parameters. Section 4 describes the results of the forecast rationality test and the estimates of the asymmetry parameters in the forecast loss function. Section 5 discusses the implications of the previous section's findings for analyzing monetary policy. Section 6 concludes.

2 Asymmetric Non-Separable Loss

Most models of forecaster behavior assume that point predictions are generated using a loss function that increases with squared forecast errors of each variable.[4] A simple squared loss function is directionally symmetric, separable across the forecasted variables, and time invariant.

Symmetry implies that positive forecast errors are equally as costly as negative forecast errors of the same magnitude. Elliott et al. (2005) showed that the symmetry assumption can be relaxed by including a (vector of) parameter(s) that governs the degree of directional asymmetry.[5] Forecasts generated from directionally asymmetric loss functions will be biased, as the forecaster prefers to either systematically underpredict or overpredict.

Separability implies that the costs of forecast errors for one variable do not depend on the forecast errors for others. This feature is clearly undesirable if we want to allow the Fed's objectives in forecasting, say, inflation to depend on its ability to forecast output. Komunjer and Owyang (2012) generalized the asymmetric loss introduced by Elliott et al. (2005) to allow for interactions between forecast errors for different variables.[6] For example, underpredicting inflation may be less costly if output growth ends up higher than expected and more costly if output growth ends up

[4] Altering the loss function does not change the predictive density. Rather, the loss function selects the "point value" of the forecast used in, for example, policymaking. As an example, if the loss function is the mean squared error, the point value of the forecast is the mean; If the loss function is the mean absolute error, the point value of the forecast is the median.

[5] Other studies that studied forecasting under asymmetric loss include Granger (1969) and Christoffersen and Diebold (1997).

[6] Using individual Blue Chip forecasts, Komunjer and Owyang (2012) found that the forecasters' loss was increased with unexpectedly worse joint economic outcomes, i.e., lower-than-expected output growth, looser-than-expected monetary policy, and higher-than-expected inflation.

lower than expected.

Finally, some studies have argued that forecasters' behavior may change over time.[7] While Capistrán (2008) focused on how the loss function varies across Fed chairmen, it may also be possible that the cost of some forecast errors depends jointly on the state of the economy.

In this paper, we consider a vector of Federal Reserve forecasts under a general loss function that allows for asymmetry and nonseparability. The next section outlines the forecaster's loss function, provides a simple bivariate example to motivate the loss, and shows how the loss function can be used to evaluate forecast rationalizability.

2.1 The Loss Function

We adopt the multivariate, non-separable, asymmetric loss framework of Komunjer and Owyang (2012) that nests both separability and symmetry. The forecaster (in this case, the Fed) attempts to predict future values of an n vector of macroeconomic variables, $\mathbf{y}_t$. Define $\mathbf{f}_{t+s,t}$ as the s-period-ahead forecast of $\mathbf{y}_{t+s}$ computed using information available at time t. The forecaster's information set $\mathcal{F}_t$ may include lagged values of $\mathbf{y}_t$ in addition to other covariates used to predict $\mathbf{y}_{t+s}$. Hereafter, we focus on one-period-ahead forecasts and set $s = 1$. Define $\mathbf{e}_{t+1}$ as the n vector of forecast errors for $\mathbf{y}_{t+1}$, i.e., $\mathbf{e}_{t+1} \equiv (\mathbf{y}_{t+1} - \mathbf{f}_{t+1,t})$. The forecaster's problem is to construct a prediction which, conditional on information known at time t, minimizes the expected loss $E\left[L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}) \mid \mathcal{F}_t\right]$ with:

$$L_p\left(\boldsymbol{\tau}, \mathbf{e}\right) \equiv \left(\|\mathbf{e}\|_p + \boldsymbol{\tau}'\mathbf{e}\right) \|\mathbf{e}\|_p^{p-1}, \tag{1}$$

where $\mathbf{e}$ is an n vector of forecast errors, and $\|\mathbf{e}\|_p$ denotes its p-norm.[8]

The loss function (1) has $n+1$ parameters: a shape parameter $p \in [1, \infty)$ and an asymmetry parameter $n-$vector $\boldsymbol{\tau}$ that belongs to the unit ball in $\mathbb{R}^n$ equipped with the norm $\|\cdot\|_q$, where $1/q + 1/p = 1$, i.e., $\boldsymbol{\tau} \in \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_q \leq 1\}$. The components $(\tau_1, \ldots, \tau_n)$ of the $n-$vector $\boldsymbol{\tau}$ govern the degree and direction of asymmetry associated with each of the forecasted variables. For $\boldsymbol{\tau} = \mathbf{0}$, there is no asymmetry: $L_p\left(\boldsymbol{\tau}, \mathbf{e}\right)$ places equal weights on forecast errors in either direction. Values of τ_i greater (less) than zero indicate greater loss for positive (negative) forecast errors in the

[7] For example, Rossi and Sekhposyan (2011) and Croushore (2012) find that the results of rationality tests can vary depending on the subsample used.

[8] Komunjer and Owyang (2012) show that (i) $L_p\left(\boldsymbol{\tau}, \cdot\right)$ is continuous and nonnegative on $\mathbb{R}^n$; (ii) $L_p\left(\boldsymbol{\tau}, \mathbf{e}\right) = 0$ if and only if $\mathbf{e} = \mathbf{0}$ and $\lim_{\|\mathbf{e}\|_p \to \infty} L_p\left(\boldsymbol{\tau}, \mathbf{e}\right) = \infty$; (iii) $L_p(\boldsymbol{\tau}, \cdot)$ is convex on $\mathbb{R}^n$.

ith element of $\mathbf{y}_{t+1}$, implying greater loss for underprediction (overprediction). This means that, conditional on the information at time t, the distribution of forecast errors for that component will have a nonzero mean. The larger $|\tau_i|$ is, the greater the loss for a similar sized error and the more the distribution of the corresponding forecast error will be biased away from zero. In addition to governing the asymmetry for the single variable i, in the nonseparable case, the parameter τ_i also determines the extent to which the interaction between the forecast errors affect the loss.

When the shape parameter $p = 1$, the loss function collapses to a multivariate tick loss sometimes used in quantile estimation.[9] In the univariate case, this flexible loss family includes: (i) the squared loss function, $L_2(0, e) = e^2$ and (ii) the absolute deviation loss function, $L_1(0, e) = |e|$, as well as their asymmetric counterparts obtained when $\tau \neq 0$, which are called (iii) the lin-lin loss, $L_1(\tau, e)$ and (iv) the quad-quad loss, $L_2(\tau, e)$. For example, the latter case of (univariate) quad-quad loss has the form

$$L_2(\tau, e) = (1 + \operatorname{sgn}(e)\,\tau)\, e^2, \tag{2}$$

where sgn $(\bullet)$ is the sign function. In the case of separability—where each forecasted variable affects the loss independent of the performance of other variables, the total loss can be computed as the sum (or weighted sum) of the univariate losses for each variable. While this obtains a form of multivariate loss, separable loss excludes the possibility that the forecast errors interact.

The forecaster's problem outlined above does not depend on the specific model used to generate the forecasts. The forecast itself can be a product of a single model, the average of several models, and/or include judgment or ad factors. As argued above, one can think of the loss function as affecting the forecaster's point estimates of the coefficients of whatever model chosen. This means that the model that the forecaster uses does not affect the econometrician's estimates of the forecaster's loss function parameters. Also, these estimates do not depend on the scheme—rolling or recursive—that the forecaster uses to construct the time series of forecasts.

[9] Lack of a natural basis for ordering multivariate data is a major obstacle in extending the notion of quantiles to multiple dimensions. The approach proposed by Chaudhuri (1996), for example, extends the tick loss function used to define univariate quantiles (see, e.g., Koenker and Bassett Jr. (1978)) to multidimensions. The multivariate tick loss $L_1(\tau, \cdot)$ leads to Chaudhuri (1996) geometric quantiles for multivariate data.

2.2 An Illustrative Example

The consequences of nonseparable asymmetric loss can be made more apparent by describing the bivariate case, $n = 2$. For simplicity of exposition, also assume $p = 2$. In this case, the loss function (1) can be rewritten as

$$L_2(\boldsymbol{\tau}, \mathbf{e}) = e_1^2 + e_2^2 + (\tau_1 e_1 + \tau_2 e_2)\left(e_1^2 + e_2^2\right)^{1/2}. \tag{3}$$

The shape of the iso-loss curves—representing combinations of forecast errors corresponding to constant loss—is determined by the parameters τ_1 and τ_2. The separable bivariate analogue of (3) can be constructed from the sum of univariate quad-quad losses:

$$L_2(\boldsymbol{\tau}, \mathbf{e}) = e_1^2 + \tau_1 \operatorname{sgn}(e_1)\, e_1 + e_2^2 + \tau_2 \operatorname{sgn}(e_2)\, e_2. \tag{4}$$

For general values of τ_1 and τ_2, the loss (3) is nonseparable in the forecast errors e_1 and e_2. This nonseparability is an important feature as it allows the forecast errors of one variable to affect the forecasts in the other variable. In particular, even if loss is directionally symmetric in the first variable (i.e., $\tau_1 = 0$), asymmetry in the second variable (i.e., $\tau_2 \neq 0$) can produce bias in the forecasts of y_1. This effect is revealed in the third term of (3) when $\tau_1 = 0$. The loss induced by e_1 when the loss is nonseparable is $e_1^2 + (\tau_2 e_2)\left(e_1^2 + e_2^2\right)^{1/2}$, which depends on both the magnitude and direction of e_2.[10] In the separable case, the losses for the two variables are independent by definition. This cross-dependence in the nonseparable case suggests an alternative explanation of the biased forecasts previously documented in the literature: It is not that the forecaster's loss is asymmetric in the the first variable; rather, the loss is symmetric in e_1 but its magnitude depends on the error committed in forecasting the second variable. In this sense, our loss (3) captures the aforementioned effect that underpredicting inflation is less costly when output is higher than expected and more costly when output is lower than expected.

As an example, suppose that a firm is forecasting two variables: market demand and its competitor's supply. Suppose also that the cost of storing product is greater than the sale price, creating an asymmetry. In this case, the forecast errors of the two variables could interact in the forecaster's

[10] Note, however, that if $\tau_1 = 0$, the direction of e_1 does not enter the loss.

loss function. Overestimating demand and underestimating the competitor's supply would be worse than any other combination of directional forecast errors because the former requires the firm to pay storage costs for the product.

Figure 1 shows the iso-loss curves for several different parameterizations. When $\tau_1 = \tau_2 = 0$, the loss $L_2(\boldsymbol{\tau}, \mathbf{e})$ is symmetric and the iso-loss curves are perfectly circular. If either $\tau_1 \neq 0$ or $\tau_2 \neq 0$, the iso-loss curves are warped in the direction of the asymmetry. The directions in which the iso-loss curve is the closest to the origin are the ones in which the loss increases the most.

We can also examine the properties of the forecasts that would be produced from various parameterizations of the loss function.[11] Suppose that $\mathbf{y}_t$ is generated from a VAR(1):

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}\mathbf{y}_{t-1} + \varepsilon_t, \tag{5}$$

where $\boldsymbol{\varepsilon}_t$ is iid multivariate normal with zero mean and covariance matrix Σ.[12] Based on the loss function, the one-period-ahead forecast is $\hat{\mathbf{f}}_{t+1,t} = \hat{\mathbf{c}} + \hat{\mathbf{A}}\mathbf{y}_t$, where

$$(\hat{\mathbf{c}}, \hat{\mathbf{A}}) \equiv \arg\min_{(\mathbf{c}, \mathbf{A})} P^{-1} \sum_{t=1}^{P} L_2(\boldsymbol{\tau}, \mathbf{y}_{t+1} - \mathbf{c} - \mathbf{A}\mathbf{y}_t), \tag{6}$$

which minimizes the expected value of the loss conditional on the data and a correctly-specified VAR(1). We then construct $P = 250$ periods of bivariate forecasts for given sets of asymmetry parameters using the same generated data.

The joint distribution of the resulting time series of forecast errors is shown in Figure 2. The directional bias of the forecast errors is evident from the plots.

2.3 Implications for Rationality Testing

Differences in the loss function also have implications for testing rationality. Past studies have sought to test the rational expectations hypothesis by evaluating private sector forecasts. Using the earliest methods (e.g., the Theil-Mincer-Zarnowitz regressions), a forecast was deemed rational if the forecast errors were mean zero and uncorrelated with information available at the time the

[11] See also Weiss (1996) for an earlier discussion of optimal forecasts produced under the forecast cost function.

[12] To generate the data, we simulate $T = R + P - 1$ periods of data from the VAR(1) after discarding the first 1000 periods to remove any initial values effects. The forecaster uses a rolling window of size $R = 100$ to construct $P = 250$ one-period-ahead forecasts.

forecast was made.[13] Many, if not most, of these studies found evidence against rationality. These methods, however, implicitly assume that the forecaster's loss function is symmetric (quadratic, in most cases).

More recently, Elliott et al. (2005, 2008) showed that private sector forecasts can be rationalizable under asymmetric loss. Unlike squared or absolute error loss, asymmetric loss assigns different penalties depending on whether the realization was above or below the forecast. As an example, Elliott et al. (2005) found that private sector forecasts could be rationalizable if forecasters were attaching more loss to overpredicting output growth than underpredicting.

3 Data and Estimation

3.1 Data

Our dataset contains three elements: the forecasts, the realizations, and the instruments used to test rationality. The full sample period is 1967:03 to 2011:12. The forecast data are for the output growth rate, the inflation rate, and the unemployment rate taken from the Greenbook.[14] The Greenbook forecasts are publicly available at a 5-year lag, vary in frequency over the sample period, and are obtained from the Philly Fed database.[15] In this paper, we consider the one-quarter-ahead forecasts. Longer-horizon Greenbook forecasts are available. However, the forecasts are conditional on an assumed (subjective) interest rate path. Over the course of the forecast exercise, the suggested path may be revised if the path is unlikely to be realized. Longer-horizon forecasts would be subject to deviations from the conditioning path. When we consider longer forecast horizons—e.g., four-quarters-ahead, we find that the loss becomes increasingly symmetric in inflation and growth, and slightly more asymmetric for unemployment. We observe the same qualitative movements when considering a separable loss. However, the movements toward symmetry (asymmetry) in growth and inflation (in unemployment) are stronger than in our benchmark non-separable loss.

[13] See Theil (1958), Mincer and Zarnowitz (1969), Figlewski and Wachtel (1981), Mishkin (1981), Zarnowitz (1985), and Keane and Runkle (1990).

[14] Greenbook forecasts are prepared by the Board of Governors staff and distributed a few days before each FOMC meeting. Forecasts are constructed using both large scale econometric models and subjective assessments. The forecasts do not necessarily reflect the opinions of the FOMC or the Board of Governors, as they are calculated before the FOMC makes a decisions on policy.

[15] These data are available here: https://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data

At the beginning of the sample, the Greenbook forecasts were available monthly; after 1981, the Greenbook was produced only for FOMC meetings. Thus, prior to 1981, we have twelve monthly vectors of observations per year; after 1981, we have eight or nine observations per year at irregular intervals. The maximum forecast horizon varies across Greenbooks in particular for the earlier dates in our sample. In addition to changes in frequency, the Greenbook changes the forecasted output growth variable in 1992 from GNP growth to GDP growth. In order to remain consistent, when the Greenbook changes the forecasted variable, we change the realization accordingly.

In constructing the output growth and inflation forecast errors, we use real GNP and its implicit deflator with fixed weights up to 1991. From December 11, 1991 to 1995 we use real GDP growth (fixed weights) and its implicit GDP deflator. The most recent date for this series available in ALFRED corresponds to the end of 1995. From the end of 1995, we use real GDP growth (chained weighted) and inflation measures based on chain-weighted implicit deflators.

The proper choice of the realization data is based on one's beliefs about the veracity of data revisions. One could believe, for example, that the intent of the forecaster is to predict the value released in real time. That is, the forecaster tries to predict the value of, say, output growth for the first quarter of 1979 which is released in April 1979. On the other hand, one could argue that the initial release of the data are poor estimates and that the revisions that occur over time are closer to the truth. If the forecaster's intent is to predict this latter value, one should use the most recent vintage of the data. Similar arguments can be made for any intermediate vintage under the assumption that significant amounts of data revisions are unpredictable and outside the scope of most agents' forecasting problems. Three common approaches are taken in the literature. The first one is to use as realizations the first release of the data; the second option is to use the one-year-later revision of the data; the third approach is to use the latest vintage. We report results using the latest available vintage as the realization.[16]

Finally, the instruments used to test rationality are one lag of each of the forecasted series, available at the time the forecast is released.[17] Ideally, we would like to instrument with the

[16] We checked the robustness of the results using the first release and found qualitative similarities. When we construct forecast errors using real-time data (first release) similarly to Capistrán (2008), we find no changes in the estimated directional asymmetry for growth and unemployment. The asymmetry parameter for inflation is close to 0 (symmetry) in both estimations, but it changes signs: negative in our benchmark estimation and positive when using first-release realizations.These results are available upon request.

[17] In principle, one could use many lags. However, Komunjer and Owyang (2012) noted that this could lead to the common "many instruments problem" and result in size distortions of the rationality test.

information available at the time the forecast is created, but unfortunately, we cannot judge when exactly the forecast is created. We know only the time at which the Greenbook was released. We assume that any data released in the previous month was available to the forecaster. Because forecasters would not have revisions available at the time they made the forecasts, we use the previous month's vintage of the forecasted variables in our instrument set.

3.2 Estimation and Rationality Testing

Komunjer and Owyang (2012) show that the asymmetry parameters in the loss function can be estimated from the forecast errors using GMM. Using Monte Carlo evidence, they argue that the shape parameter requires a very long time series for inference—much longer than we have for the Greenbook forecasts; hence, they suggest calibrating p. Komunjer and Owyang (2012) show that the asymmetry parameters are consistently estimated for different values of p; thus, results using alternative loss functions should be similar. We report results for the shape parameter calibrated to 2, i.e., $p = 2$.[18] The estimation chooses the value of the asymmetry parameter that maximizes the GMM objective function derived from the first-order condition for the non-separable asymmetric loss.

The resulting value of the GMM objective function is used to construct the $J-$statistic for use in the rationality test. The test employs the standard test for overidentification to determine whether the forecasts are rationalizable, conditional on the instrument set (see Komunjer and Owyang, 2012, for details). For the $J-$test, the number of degrees of freedom are equal to the number of instruments minus one times the number of series forecasted. In our application, we use three instruments per series; thus, the test statistic is distributed χ^2 with 6 degrees of freedom.

3.3 Testing for State-Dependence

In order to capture potential state-dependence, we allow the forecaster's asymmetry parameter $\boldsymbol{\tau}$ to change. Let P be the sample size over which the forecast errors are observed. Specifically, we focus on the case in which $\boldsymbol{\tau}$ can switch between two different values, $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$, at a known break

[18] We tested a few different values of p ($p = 1.5$ and $p = 2.5$) and found no appreciable differences in the results. Thse results are available upon request.

time, $\widetilde{t}$.[19] To test for state dependence, we construct a Wald test of the no-change hypothesis,

$$H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2,$$

against the switching alternative,

$$H_1 : \quad \boldsymbol{\tau} = \begin{cases} \boldsymbol{\tau}_1, & \text{for } t < \widetilde{t} \\ \boldsymbol{\tau}_2, & \text{for } t \geq \widetilde{t} \end{cases}.$$

In describe the proposed test, the idea is then to use a partial-sample GMM estimator that uses the data $t < \widetilde{t}$ to estimate $\boldsymbol{\tau}_1$ and the data in $t \geq \widetilde{t}$ to estimate $\boldsymbol{\tau}_2$. Call $\hat{\boldsymbol{\tau}}_{1P}$ and $\hat{\boldsymbol{\tau}}_{2P}$ the partial-sample GMM estimators of $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$, respectively. The Wald statistic for testing H_0 against H_1 then takes the form

$$W_P \equiv P\left(\hat{\boldsymbol{\tau}}_{1P} - \hat{\boldsymbol{\tau}}_{2P}\right)' \left(\frac{1}{\pi}\left(\hat{\mathbf{B}}'_P \hat{\mathbf{S}}_{1P}^{-1} \hat{\mathbf{B}}_P\right)^{-1} + \frac{1}{1-\pi}\left(\hat{\mathbf{B}}'_P \hat{\mathbf{S}}_{2P}^{-1} \hat{\mathbf{B}}_P\right)^{-1}\right)^{-1} \left(\hat{\boldsymbol{\tau}}_{1P} - \hat{\boldsymbol{\tau}}_{2P}\right),$$

where $\pi \in [0,1]$ is the limit of the ratio $\frac{P_1}{P_1+P_2}$ when the sizes P_1 and P_2 of the partial samples go to infinity, the estimators $\hat{\mathbf{S}}_{1P}$ and $\hat{\mathbf{S}}_{2P}$ are constructed using the data in D_1 and D_2, respectively, while the estimator $\hat{\mathbf{B}}_P$ is constructed using the entire sample, i.e.,

$$\begin{aligned}
\hat{\mathbf{B}}_P &\equiv \frac{1}{T}\sum_{t=R}^{T}\left[\|\hat{\mathbf{e}}_{t+1}\|_p^{p-1}\left(\mathbf{I}_n \otimes \mathbf{x}_t\right) + (p-1)\|\hat{\mathbf{e}}_{t+1}\|_p^{-1}\left(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t\right)\hat{\mathbf{e}}'_{1,t+1}\right], \\
\hat{\mathbf{S}}_{1P} &\equiv \frac{1}{\pi T}\sum_{t \in D_1}\left[\left(\hat{\mathbf{M}}_{1,t+1}\hat{\mathbf{M}}'_{1,t+1}\right) \otimes \left(\mathbf{x}_t \mathbf{x}'_t\right)\right], \\
\hat{\mathbf{M}}_{1,t+1} &\equiv \left(p\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) + \hat{\boldsymbol{\tau}}_{1P}\|\hat{\mathbf{e}}_{t+1}\|_p^{p-1} + (p-1)\hat{\boldsymbol{\tau}}'_{1P}\hat{\mathbf{e}}_{t+1}\|\hat{\mathbf{e}}_{t+1}\|_p^{-1}\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1})\right),
\end{aligned}$$

with analogous definitions for $\hat{\mathbf{S}}_{2P}$ and $\hat{\mathbf{M}}_{2,t+1}$, where T is the last period used to evaluate the Wald statistic. Note that we can use the entire sample to construct an estimator for $\mathbf{B}$ because the latter does not depend on $\boldsymbol{\tau}$.[20] Under the null of no structural breaks, the partial-sample GMM

[19] Our framework supports tests of multiple switches between two asymmetry parameter values (e.g., asymmetry across the business cycle) or multiple breaks with more than two asymmetry parameters. The former is a simple extension of the current framework and is discussed in general in the appendix. The latter can be thought of as iteratively adding an extra break to a subsample.

[20] Technical details for the test can be found in the Appendix.

estimator $(\hat{\boldsymbol{\tau}}_{1P}, \hat{\boldsymbol{\tau}}_{2P})$ has a known asymptotic distribution, and the Wald statistic is asymptotically $\chi^2(n)$ distributed, where n is the number of forecast variables. Appendix B details the assumptions needed for the asymptotic results to hold.

4 Empirical Results

In this section, we assess the directional asymmetry in the Fed's forecast loss function, estimated using data from the Greenbook forecast. In each case, the estimated $\boldsymbol{\tau}$ can be interpreted as the (smallest) degree of asymmetry most consistent with rationality, assuming the forecast can be rationalized; larger magnitudes of $\boldsymbol{\tau}$ suggest more asymmetric preferences. In evaluating the results, it may be helpful to keep in mind that $\tau_n > 0$ indicates greater loss for positive forecast errors (i.e., greater loss for underprediction) and that $\tau_n < 0$ indicates greater loss for negative forecast errors (i.e., greater loss for overprediction). Values of $\boldsymbol{\tau}$ that are not statistically different from zero imply symmetric loss.

4.1 Benchmark Full Sample Results

As a benchmark, we first estimate a set of (constant) asymmetry parameters for the full sample of Greenbook forecasts (1967:03 to 2011:12) using both separable loss—which assumes independent losses for all of the variables—and non-separable loss—which assumes that the directional losses interact. We obtain the separable loss function parameters by estimating univariate versions of the loss function, equation (2).

The bottom row of Table 1 shows the values of the $J-$statistic described in Komunjer and Owyang (2012) used to test rationalizability. Consistent with previous studies of both Fed and private sector forecasts, the Greenbook forecasts appear to be rationalizable if one takes into account potential directional asymmetry. This result obtains regardless of whether we assume that the loss function is separable or nonseparable; separability only affects the estimates of the minimum degree of asymmetry required to obtain rationality.[21]

The top row of Table 1 compares the estimates obtained with the nonseparable (first three

[21] One might be concerned that using a flexible loss results in an "automatic" finding of rationality. Komunjer and Owyang (2012), however, found that some private sector forecasters can be found to be unrationalizeable even under asymmetric loss.

columns) and separable (last three columns) loss functions for the full sample period. Consistent with the full sample result in Capistrán (2008), the Greenbook forecasts for inflation appear unbiased (and, hence, symmetric). This result is invariant to the assumption of separable versus nonseparable loss. The asymmetry parameter most consistent with rationality for output growth exhibits a substantial degree of asymmetry. In particular, overpredictions of output growth are associate with greater losses. The interpretation of this result is straightforward: weaker economic performance is costlier to the Fed than a surprise boom of the same magnitude.

The difference between separable and non-separable loss is highlighted by the results for the unemployment rate. According to our estimates if the loss function is assumed to be separable, the Fed prefers to overpredict the unemployment rate. This results is consistent with the previous result for output growth, where surprise economic weakness is more costly than a surprise boom. However, if the magnitude and direction of the inflation and output growth forecast errors are taken into account, the loss for unemployment is essentially symmetric—the unemployment forecast errors are conditionally unbiased. To quantify these results, the loss for an equivalent underprediction of unemployment would be viewed as ten times larger under separable loss than under nonseparable loss.

This result could imply that policymaking is conducted under the belief in either a Phillips curve or Okun-type relationship that relates these variables. In either case, once asymmetry in inflation and output growth are accounted for, unemployment forecast errors appear symmetric.[22] We revisit this issue below.

4.2 Changes in the Fed's Behavior

The empirical literature on the behavior of U.S. monetary policy is rife with instances of instability attributed to any number of different root causes: Changes in the Fed leadership, differences in the Fed's operating procedure, or variation in the volatility of macroeconomic aggregates (i.e., the Great Moderation) could be thought to alter the Fed's forecast behavior. For example, Capistrán (2008) examines whether the forecast behavior of the Fed has changed over time, focusing primarily

[22] While the parameter remains statistically different from zero, the value is small, implying that the relative differences in the directional losses are very small.

on the appointment of Paul Volcker in 1979.[23]

On the other hand, the Great Moderation has led forecasting—in particular, inflation forecasting—to change. The change in inflation's forecastability subsequent to the onset of the Great Moderation was highlighted in Atkeson and Ohanian (2001), who argued that simple random-walk forecasts outperform Phillips curve models during the Great Moderation period, and Stock and Watson (2007), who argue for an integrated moving-average forecast of inflation given the variation in trends over the U.S. post-War period.

4.2.1 Break Tests

The aforementioned studies suggest two possible—perhaps not mutually—exclusive break dates. Given the potential differences in the interpretation of the loss function asymmetry parameter shift occurring at various times, we tested three break scenarios using the test proposed above: (1) a single break at the Volcker appointment (October, 1979), (2) a single break at the Great Moderation assumed to occur in January 1984 (see McConnell and Perez-Quiros (2000)), and (3) a break at the Great Moderation assuming a pre-existing break at the Volcker appointment.

Table 2 contains the results of our break tests. The left panel shows the results of the break tests for nonseparable loss and the right panel shows the results for separable loss. The first row of each panel shows the Wald statistic and the associated p-value (in parentheses) for a joint break in the vector of asymmetry parameters. The first column of each panel shows the results of a test for a single break at the Volcker appointment. The second column of each panel shows the results of a test for a single break at the onset of the Great Moderation. The third column of each panel shows the results of a test for a second break at the Great Moderation assuming a break at the Volcker appointment. The bottom three rows contain the results for tests considering breaks in each individual data series.[24] These experiments suggest which of the forecast error series appear to be the most affected by changes in Fed behavior.

We find strong evidence of a break at the time of the Volcker appointment in both the joint model and in the loss for individual series. The exception is the unemployment rate. We also find

[23] Capistrán argues that overprediction is more costly in the pre-Volcker sample, while underprediction is more costly in the post-Volcker sample. Capistrán's univariate analysis is equivalent to isolating the inflation parameter in a separable loss function. In this sense, our loss function can be thought of as nesting the baseline model in Capistrán.

[24] For separable loss, this is equivalent to testing for a break in a univariate environment. For nonseparable loss, we are testing for a break in the loss for one series but not the other two.

evidence of a second break in the loss using the joint model occurring at the time of the Great Moderation after the sample is split at the Volcker appointment. The break results suggest three subsamples for the forecast loss: (1) the pre-Volcker period, (2) the Volcker disinflation—the time between the Volcker's appointment and the beginning of the Great Moderation, and (3) the Great Moderation period. Results are consistent across separable and nonseparable loss except for the case testing two breaks in the unemployment rate. In the separable case, we find some evidence of a break in the loss for unemployment at the Great Moderation, given a break at the Volcker appointment. If we consider the same scenario under nonseparable loss, we reject a the Great Moderation break for the post-Volcker subsample.

Whether the cause is a decline in macroeconomic volatility (as in Atkeson and Ohanian (2001) and Stock and Watson (2007)) or an increase in the central bank's anti-inflation credibility (as in Capistrán (2008)) in the wake of the Volcker disinflation, it seems sensible to analyze independently the forecasts before the Volcker appointment and subsequent to the onset of the Great Moderation. In addition to including two breaks, we allow the forecast errors to interact in the loss function, taking into account the possibility of the forecaster's belief in Phillips curve-type macroeconomic relationships.

4.2.2 Asymmetry Parameters

The results reported here are for two joint breaks in the vector of asymmetry parameters creating three subperiods. Table 3 contains the estimates of the asymmetry parameters for nonseparable (top panel) and separable (bottom panel) loss for the pre-Volcker period (right column) and the Great Moderation period (left column). We omit reporting of the intervening period (the Volcker disinflation) as the period is short (about five years) and has extremely volatile inflation and interest rates. The first three rows of each panel contain the asymmetry parameters and their standard errors (in parentheses). The last row of each panel contains the $J-$statistic used to test subsample rationality.

Consistent with previous studies, we find that the Fed prefers underpredicting inflation prior to Volcker. However, in contrast with Capistrán (2008), who argued the Fed prefers overpredicting inflation after Volcker is appointed, we find that, once the second break is introduced, the Fed's loss for inflation subsequent to the Volcker disinflation is symmetric. This result obtains regardless of whether one assumes separable or nonseparable loss. We conclude that the result that the Fed prefers overprediction after Volcker is a product of the highly volatile inflation rate occurring during the Volcker disinflation.[25]

The asymmetry parameters for output change in a manner opposite to inflation. While the loss for output is symmetric prior to Volcker, the Fed prefers underpredicting output in the Great Moderation period. Finally, consistent with the full sample results, unemployment rate forecasts appear to be asymmetric when analyzed in isolation (i.e., under separable loss) but appear essentially symmetric if analyzed together with other forecasts (i.e., under nonseparable loss). In either case, the Fed appears to prefer overpredictions of unemployment.

Interpreting individual loss is relatively straightforward: the sign determines the direction of the asymmetry and the magnitude determines the extent of the asymmetry.[26] But how do we interpret the asymmetries relative to each other in the nonseparable case? For example, what does it mean for the Fed to switch from being symmetric for growth and underpredicting inflation in the pre-Volcker period but underpredicting growth and being symmetric about inflation in the Great Moderation period.

One might interpret the shift in the Fed's loss function as a change from more activist policy to more conservative policy. Suppose the Fed believes there to be a relationship between inflation, output, and a third variable: the policy instrument. Before the Volcker appointment, the Fed's objective was to accurately predict output with less regard for how policy would affect inflation. After the onset of the Great Moderation, the Fed's objective was shifted to accurately predicting inflation while avoiding optimism about the rate of economic growth. This implies both a strengthening the inflation objective and an unwillingness to experience negative surprises on output growth.

4.2.3 Complementarities and Their Interpretation[27]

In addition to the shifts in the asymmetry parameters across subsamples, we found that the losses for unemployment forecast errors are conditionally symmetric. This suggests that policy makers believe

[25] During this period, the inflation rate was falling consistently. If it fell faster than anticipated, this would explain the isolated period of overpredictions.

[26] For example, under separable loss and $p = 2$ with $\tau = 0.25$. In this case, an underprediction produces twice the loss as an overprediction of the same magnitude. Under nonseparable loss, however, the loss depends on the forecast erros of the other variables.

[27] The derivation for the complementarities are shown in the Appendix.

some complementarities exist between unemployment and other macroeconomic variables and leads us to a number of questions. Which variable—output growth or inflation—is more important in determining the loss in unemployment rate forecasts? Are these relationships stable over time? What do they mean for interpreting forecaster behavior? In order to answer these questions and possibly differentiate between possible explanations, we construct a measure of the complementarity between two variables, γ_{ij}, defined as the second cross-derivative of the loss function with respect to variables i and j evaluated at the optimum.[28] For a given value of p, a consistent estimator $\hat{\gamma}_{ij}$ of γ_{ij} is simply obtained by plugging in the values of a consistent estimator $\hat{\boldsymbol{\tau}}$ of $\boldsymbol{\tau}$, and by replacing the expectations with their sample counterparts. In particular, the full-sample estimator $\hat{\gamma}_{ij}$ of γ_{ij} is obtained as:

$$\hat{\gamma}_{ij} = \frac{p-1}{T} \sum_{t=R}^{T} \left[\frac{\hat{\tau}_i \operatorname{sgn}(\hat{e}_{jt})|\hat{e}_{jt}|^{p-1} + \hat{\tau}_j \operatorname{sgn}(\hat{e}_{it})|\hat{e}_{it}|^{p-1}}{\|\hat{\mathbf{e}}_{\mathbf{t}}\|_p} - \hat{\boldsymbol{\tau}}'\hat{\mathbf{e}}_{\mathbf{t}} \frac{\operatorname{sgn}(\hat{e}_{it}) \operatorname{sgn}(\hat{e}_{jt})|\hat{e}_{it}|^{p-1}|\hat{e}_{jt}|^{p-1}}{\|\hat{\mathbf{e}}_{\mathbf{t}}\|_p^{p+1}} \right], \tag{7}$$

where R corresponds to the initial forecasting period. The complementarity defines the loss associated with a simultaneous change in the forecast errors in two variables. The sign of the complementarity suggests how the errors in each variables interact: positive complementarities imply higher losses when the variables have the same signed forecast errors. By definition, the complementarities for separable loss are identically zero.

We can interpret the complementarity as suggestive of whether FOMC members believe in a relationship between the two variables. If policymakers believe that there is a relationship between two variables, the sign of the complementarity should be consistent with unambiguously better or worse outcomes. For example, if policymakers believe that underpredicting inflation generates worse outcomes when also underpredicting unemployment rates, the complementarity for these variables should be positive.

Table 4 shows the values of the bivariate complentarities. The top row shows the full sample complementarities while the bottom two rows show subsample complementarities. Based on the argument above, we can surmise that the Fed's belief in the relationship between errors in inflation and output growth forecasts has been stable in time. The positive sign of the inflation/output

[28] To compute the complementarities, we estimate all combinations of bivariate losses. We construct the bivariate complementarities because trivariate complementarities are both difficult to compute and to interpret.

complementarity implies that loss increases when the forecast errors in inflation and output growth both increase or both decrease—i.e., when the direction of change in the forecast error is the same.

We interpret the positive sign in the complementarity between inflation and unemployment rates as suggestive of the policy makers believes in a Phillips curve relationship—i.e., their loss increases when both inflation and unemployment forecast errors increase or decrease. However, this belief appears unstable over time. The positive sign of the inflation/unemployment complementarity is mostly driven by the beliefs of policy makers in the pre-Volcker era. The sign of the complementarity shifts to negative in the post-Great Moderation period, suggesting that the loss rises when the forecast error of one variable goes up and the other goes down, consistent with a weaker belief from the policy maker in a Phillips curve relationship. This finding is consistent with empirical studies of effect of FOMC member biases on inflation forecasts, Malmendier et al. (2017).

5 Implications for How We Interpret Monetary Policy

Our results may have additional implications for the historical analysis of monetary policy. Clarida et al. (2000) argued that, prior to Volcker's appointment, the Fed's policy rule put too little emphasis on stabilizing inflation. Subsequently, under Volcker and Greenspan, the Fed responded to expected inflation more aggressively, resulting in the end of the Great Inflation. Orphanides (2004), using real-time data and Greenbook forecasts, finds that the policy rule was essentially stable since the 1960s except for the output response. Thus, Orphanides concludes that the period of high inflation was the result of real-time data leading to inaccurate forecasts rather than a policy not aggressive enough in fighting inflation.

If the Fed's forecasting behavior is indeed related to its perception of policy risk, as Ellison and Sargent (2012) suggests, the asymmetries may imply that the Fed perceived the risk of being above the inflation objective differently under the two regimes. Capistrán's result tends to support Clarida et al. (2000), claiming that the Fed became more pro-active about maintaining price stability simultaneously with an increased loss for underpredicting inflation. Indeed, we also find a change in the forecast behavior for inflation consistent with this directionality; however, we do not find that the inflation forecasts are overpredictive once we account for the forecast errors of the other variables. However, (Orphanides, 2004, 2001) argue against changes in policymaker preferences, opting instead to attribute the differences in policymaker behavior to variation in the quality of the real-time data. In addition, studies in the forecasting literature suggest that the volatility reduction that occurred during the Great Moderation may have changed the overall forecastability of macroeconomic aggregates.

Could our result be consistent with Orphanides? If the pre-Volcker Fed cared about output and had a misspecified Phillips curve, the Greenbook inflation forecasts may have been consistently be low. Thus, prior to the Volcker break, the forecast errors are consistent with a model in which the policymaker believes in an inflation-output trade-off. Subsequent to the onset of the Great Moderation, an apparent inflation-output trade-off still exists but more emphasis appears to be placed on accurate inflation forecasts. In a Taylor-type policy rule, "overly optimistic" downward biased inflation forecasts can result in overly accommodative policy. The forecast biases appear even controlling for the real-time data environment, suggesting a preference-based alternative to Orphanides view from the trenches. As in Orphanides, policymakers can be forward-looking and attempting to implement optimal policy, yet still induce instability based on biases in the forecasts. In our case, however, these biases appear to be recurring, if not intentional.

6 Conclusions

We study the forecasting behavior of the Federal Reserve and attempt to understand its implications for policy design. We find evidence of systematic bias in the Greenbook forecasts of inflation and output growth used to conduct policy. Biases in the unemployment forecasts are mitigated when taken into context with the directionally-biased forecasts in the other variables.

We find an interesting pattern in the forecasts when considering subsamples prior to the Volcker appointment and subsequent to the onset of the Great Moderation. In the former subperiod, inflation forecasts are biased downward but output growth forecasts are essentially unbiased. In the latter period, the opposite is true: Inflation forecast errors are essentially mean zero and output forecasts appear conservative.

These results are broadly consistent with other studies on the change in monetary policy that occurred in the late 1970s, which has been interpreted by Orphanides as the Fed coming to terms with the difficulties in predicting real variables and by Capistrán as a change in the Fed's forecast behavior for inflation. Our results extend and refine Capistrán's in a multivariate setting, where the inflation forecasts are not taken in isolation. The change in the policy behavior observed in the literature appears to be simultaneous with a shift in the Fed's forecast behavior.

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–56.

Andrews, D. W. & Fair, R. C. (1988). Inference in nonlinear econometric models with structural change. *Review of Economics Studies*, 55(6):615–40.

Atkeson, A. & Ohanian, L. E. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis Quarterly Review*, 25(1):2–11.

Brayton, F., Roberts, J. M., & Williams, J. C. (1999). What's happened to the phillips curve?

Capistrán, C. (2008). Bias in federal reserve inflation forecasts: Is the federal reserve irrational or just cautious? *Journal of Monetary Economics*, 55(8):1415–27.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872.

Christoffersen, P. & Diebold, F. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, 13(06):808–817.

Clarida, R. J., Galí, J., & Gertler, M. L. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics*, 115(1):147–80.

Croushore, D. (2012). Forecast bias in two dimensions. University of Richmond, unpublished manuscript.

Elliott, G., Komunjer, I., & Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economics Studies*, 72(4):1107–25.

Elliott, G., Komunjer, I., & Timmermann, A. (2008). Biases in macroeconomic forecasts: Irrationality or asymmetric loss? *Journal of the European Economic Association*, 6(1):122–57.

Ellison, M. & Sargent, T. J. (2012). A defense of the fomc. *International Economic Review*, 53(4):1047–1065.

Figlewski, S. & Wachtel, P. (1981). The formation of inflationary expectations. *The Review of Economics and Statistics*, 63(1):1–10.

Granger, C. (1969). Prediction with a generalized cost of error function. *Operational Research Quarterly*, (20):199–207.

Jansen, D. W. & Kishan, R. P. (1996). An evaluation of federal reserve forecasting. *Journal of Macroeconomics*, 18(1):89–109.

Keane, M. P. & Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *American Economic Review*, 80(4):714–35.

Koenker, R. & Bassett Jr., G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Komunjer, I. & Owyang, M. T. (2012). Multivariate forecast evaluation and rationality testing. *Review of Economics and Statistics*, 94(4):1066–80.

Malmendier, U., Nagel, S., & Yan, Z. (2017). The making of hawks and doves: Inflation experiences on the fomc. Working Paper 23228, National Bureau of Economic Research.

McConnell, M. M. & Perez-Quiros, G. (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review*, 90(5):1464–1476.

Mincer, J. A. & Zarnowitz, V. (1969). The evaluation of economic forecasts. In: *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, J. A. Mincer, ed., Studies in Business Cycles, chapter 1, pages 1–46. National Bureau of Economic Research.

Mishkin, F. S. (1981). Are market forecasts rational? *American Economic Review*, 71(3):295–306.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, 91(4):964–85.

Orphanides, A. (2002). Monetary-policy rules and the Great Inflation. *American Economic Review*, 92(2):115–20.

Orphanides, A. (2004). Monetary policy rules, macroeconomic stability, and inflation: A view from the trenches. *Journal of Money, Credit, and Banking*, 36(2):151–75.

Romer, C. D. & Romer, D. H. (2000). Federal reserve information and the behavior of interest rates (digest summary). *American Economic Review*, 90(3):429–457.

Rossi, B. & Sekhposyan, T. (2011). Forecast optimality tests in the presence of instabilities. Economic Research Initiatives at Duke (ERID) Working Paper No. 109.

Sargent, T., Williams, N., & Zha, T. (2006). Shocks and government beliefs: The rise and fall of american inflation. *American Economic Review*, 96(4):1193–1224.

Sims, C. A. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity*, 2002(2):1–62.

Stock, J. H. & Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, 39(s1):3–33.

Swanson, E. T. (2004). Assessing the rationality of Federal Reserve "Greenbook" forecasts. Unpublished manuscript.

Tarullo, D. (2017). Monetary policy without a working theory of inflation. Working Paper 33, Brookings.

Theil, H. (1958). *Economic Forecasts and Policy*. North Holland Publishing Co.

Wang, Y. & Lee, T.-H. (2012). Asymmetric loss in the Greenbook and the Survey of Professional Forecasters. UC Riverside, unpublished manuscript.

Weiss, A. (1996). Estimating time series models using the relevant cost function. *Journal of Applied Econometrics*, (11):539–60.

West, K. D. (2006). Forecast evaluation. In: *Handbook of Economic Forecasting*, G. Elliott, C. W. Granger, & A. Timmermann, ed., volume 1, chapter 3, pages 99–134. Elsevier Science, North-Holland.

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

A Notation

For any scalar u, $u \in \mathbb{R}$, we let $\mathbb{1} : \mathbb{R} \to [0,1]$ be the Heaviside (or indicator) function, i.e., $\mathbb{1}(u) = 0$ if $u < 0$, $\mathbb{1}(u) = 1$ if $u > 0$, and $\mathbb{1}(0) = \frac{1}{2}$. Similarly, we use $\mathrm{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ to denote the sign function: $\mathrm{sgn}(u) = \mathbb{1}(u) - \mathbb{1}(-u) = 2\mathbb{1}(u) - 1$, and let $\delta : \mathbb{R} \to \mathbb{R}$ be the Dirac delta function. Note that the Heaviside function is the indefinite integral of the Dirac function, i.e., $\mathbb{1}(u) = \int_a^u d\delta$, where a is an arbitrary (possibly infinite) negative constant, $a \leqslant 0$. For any real function $f : \mathbb{R}^n \to \mathbb{R}$ that is continuously differentiable to order $R \geqslant 2$ on $\mathbb{R}^n$, we let $\nabla_{\mathbf{u}} f(\mathbf{u})$ denote the gradient of $f(\cdot)$ with respect to $\mathbf{u}$, $\nabla_{\mathbf{u}} f(\mathbf{u}) \equiv (\partial f(\mathbf{u})/\partial u_i, ..., \partial f(\mathbf{u})/\partial u_n)'$, and use $\Delta_{\mathbf{uu}} f(\mathbf{u})$ to denote its Hessian matrix, $\Delta_{\mathbf{uu}} f(\mathbf{u}) \equiv (\partial^2 f(\mathbf{u})/\partial u_i \partial u_j)_{1 \leqslant i,j \leqslant n}$.

For any n-vector $\mathbf{u}$, $\mathbf{u} = (u_1, ..., u_n)' \in \mathbb{R}^n$, we denote by $\|\mathbf{u}\|_p$ its l_p-norm, i.e., $\|\mathbf{u}\|_p = (|u_1|^p + ... + |u_n|^p)^{1/p}$ for $1 \leqslant p < \infty$, and $\|\mathbf{u}\|_\infty = \max_{1 \leqslant i \leqslant n}(|u_i|)$. We define the open unit ball $\mathcal{B}_p^n$ in $\mathbb{R}^n$ as $\mathcal{B}_p^n = \{u \in \mathbb{R}^n : \|\mathbf{u}\|_p < 1\}$. We denote by $\boldsymbol{\iota}_i$ an n-vector whose only non-zero component is in its ith position and it equals 1, i.e. $\boldsymbol{\iota}_1 \equiv (1, 0, 0, \ldots, 0)'$, $\boldsymbol{\iota}_2 \equiv (0, 1, 0, \ldots, 0)'$, etc...

$\boldsymbol{\nu}_p(\mathbf{u})$, $\mathbf{V}_p(\mathbf{u})$ and $\mathbf{W}_p(\mathbf{u})$ are an n-vector and two $n \times n$-diagonal matrices defined as:

$$
\begin{aligned}
\boldsymbol{\nu}_p(\mathbf{u}) &\equiv (\mathrm{sgn}(u_1)|u_1|^{p-1}, \ldots, \mathrm{sgn}(u_n)|u_n|^{p-1})' \\
\mathbf{V}_p(\mathbf{u}) &\equiv \mathrm{diag}(\delta(u_1)|u_1|^{p-1}, \ldots, \delta(u_n)|u_n|^{p-1}) \\
\mathbf{W}_p(\mathbf{u}) &\equiv \mathrm{diag}(|u_1|^{p-2}, \ldots, |u_n|^{p-2})
\end{aligned}
$$

Then

$$\nabla_{\mathbf{u}} \|\mathbf{u}\|_p = \|\mathbf{u}\|_p^{1-p} \, \boldsymbol{\nu}_p(\mathbf{u})$$

and

$$\Delta_{\mathbf{uu}} \|\mathbf{u}\|_p = \|\mathbf{u}\|_p^{1-p} \left\{ 2\mathbf{V}_p(\mathbf{u}) + (p-1) \left[\mathbf{W}_p(\mathbf{u}) - \|\mathbf{u}\|_p^{-p} \, \boldsymbol{\nu}_p(\mathbf{u}) \boldsymbol{\nu}_p'(\mathbf{u}) \right] \right\},$$

which we shall often be using in what follows.

B Estimation and Inference

B.1 Forecasting Setup

We consider the setup in which, at time t, the forecaster is interested in constructing one-step-ahead predictions of an n-vector of interest, $\mathbf{y}_{t+1}$ ($n \geqslant 1$). Letting $\mathbf{f}_{t+1,t}$ and $\mathbf{e}_{t+1} \equiv \mathbf{y}_{t+1} - \mathbf{f}_{t+1,t}$ denote the one-step-ahead forecast and the corresponding forecast error, we shall assume that the forecasts are constructed to minimize the expected n-variate loss, $L_p(\boldsymbol{\tau}_0, \cdot)$,

$$L_p(\boldsymbol{\tau}_{\mathbf{0}}, \mathbf{e}) \equiv \left(\|\mathbf{e}\|_p + \boldsymbol{\tau}_{\mathbf{0}}{}'\mathbf{e} \right) \|\mathbf{e}\|_p^{p-1}, \tag{8}$$

with unknown parameter $\boldsymbol{\tau}_0 \in \mathcal{B}_q^n$, where $1/p + 1/q = 1$ and $1 \leqslant p < \infty$ is known. Specifically, the forecaster is assumed to solve the following minimization problem:

$$\mathbf{f}^*_{t+1,t} = \arg \min_{\{\mathbf{f}_{t+1,t}\}} E\left[L_p(\boldsymbol{\tau}_0, \mathbf{y}_{t+1} - \mathbf{f}_{t+1,t}) \middle| \mathcal{F}_t \right], \tag{9}$$

where $\mathcal{F}_t$ denotes the forecaster's information set that is informative for $\mathbf{y}_{t+1}$ (e.g., lagged values of $\mathbf{y}_t$ as well as other covariates), and the loss $L_p(\boldsymbol{\tau}_0, \cdot)$ is defined as above. Komunjer and Owyang (2012) show that (9) is equivalent to the following forecast rationality condition:

$$E\left[p\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1}) + \boldsymbol{\tau}_{\mathbf{0}} \left\|\mathbf{e}^*_{t+1}\right\|_p^{p-1} + (p-1)\tau_0'\mathbf{e}^*_{t+1} \frac{\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1})}{\left\|\mathbf{e}^*_{t+1}\right\|_p} \middle| \mathcal{F}_t \right] = \mathbf{0}, a.s. - P, \tag{10}$$

where $\mathbf{e}^*_{t+1} \equiv \mathbf{y}_{t+1} - \mathbf{f}^*_{t+1,t}$ is the optimal forecast error.

We are concerned with situations in which the forecaster's asymmetry parameter $\boldsymbol{\tau}$ is allowed to change in time. We describe the case in which $\boldsymbol{\tau}$ can switch between two different values, $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$, at known switching times. This case is more general than the single break case described in the main text. However, collapsing to the one break case is straightforward. Also, multiple breaks with more than two values for $\boldsymbol{\tau}$ can be modeled by considering a single break first and then a second break in the subsample created by the first break.

Letting P be the sample size over which the forecast errors are observed, starting at $t = R$ and ending at $t = R + P - 1 \equiv T$, we denote by $t_1, t_2, \ldots, t_S$ the switching times so that

$$\boldsymbol{\tau} = \begin{cases} \boldsymbol{\tau}_1, & \text{for } t \in [R, t_1 - 1] \cup [t_2, t_3 - 1] \cup \ldots \cup [t_S, T] \\ \boldsymbol{\tau}_2, & \text{for } t \in [t_1, t_2 - 1] \cup [t_3, t_4 - 1] \cup \ldots \cup [t_{S-1}, t_S - 1] \end{cases} \tag{11}$$

if S is even. Similarly, if S is odd, then the last time intervals during which $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are observed become $[t_{S-1}, t_S - 1]$ and $[t_S, T]$, respectively.

The forecaster then uses data from 1 to R to compute the first forecast of $\mathbf{y}_{R+1}$, $\hat{\mathbf{f}}_{R+1,R}$; the estimation window is then rolled on, and data from 2 to $R+1$ is used to compute $\hat{\mathbf{f}}_{R+2,R+1}$. Up until t_1, the forecaster's degree of asymmetry equals $\boldsymbol{\tau}_1$. From t_1, i.e., starting with the forecast $\hat{\mathbf{f}}_{t_1+1,t_1}$, the forecaster's degree of asymmetry is allowed to change to $\boldsymbol{\tau}_2$ up to time t_2. The exercise continues until the end of the forecasting period, at $t = T$. We shall make the following assumption:

Assumption 1 *Let P_1 and P_2 denote the lengths of the forecasting samples corresponding to regimes 1 and 2, respectively, i.e., $P_1 \equiv (t_1 - R) + (t_3 - t_2) + \ldots + (T - t_S)$ and $P_2 \equiv (t_2 - t_1) + (t_4 - t_3) + \ldots + (t_S - t_{S-1})$ if S is even, and $P_1 \equiv (t_1 - R) + (t_3 - t_2) + \ldots + (t_S - t_{S-1})$ and $P_2 \equiv (t_2 - t_1) + (t_4 - t_3) + \ldots + (T - t_S)$ if S is odd. Then, $(R, P_1, P_2) \to \infty$ and $\frac{P_1}{P_1 + P_2} \to \pi \in [0, 1]$.*

Put in words, Assumption 1 says the following: Consider the two samples of forecast errors $\hat{\mathbf{e}}_{t+1} \equiv \mathbf{y}_{t+1} - \hat{\mathbf{f}}_{t+1,t}$: those of size P_1 constructed under $\boldsymbol{\tau}_1$, and those of size P_2 constructed under $\boldsymbol{\tau}_2$. Then, the first requirement is that both sample sizes P_1 and P_2 go to infinity. This requirement restricts the class of switching processes we can allow for, though the restriction is fairly weak. The number of switches S can be either finite or infinite and either deterministic or random. For example, if S is finite, then Assumption 1 requires that the forecaster starts the forecasting exercise with the parameter $\boldsymbol{\tau}_1$ and ends the latter with $\boldsymbol{\tau}_2$, thus ensuring that both P_1 and P_2 go to infinity. This situation would exclude situations in which the forecaster's parameter shifts from $\boldsymbol{\tau}_1$ to $\boldsymbol{\tau}_2$ and back somewhere within the forecasting sample, since in this case P_2 remains finite. With P_2 finite, it would be impossible to distinguish this case from the "no-switch" one in which the parameters remains equal to $\boldsymbol{\tau}_1$ throughout the forecasting sample. As already pointed out, the switching process can be either deterministic or random. In the latter case, however, it is necessary that the process be such that the ratio $P_1/(P_1 + P_2)$ converges to some known constant π in $[0, 1]$. The final requirement of Assumption 1 is that the length R of the first estimation window also goes to infinity, which is a standard requirement in the forecasting literature (see, e.g., West, 2006).

B.2 Estimation of Loss Function Parameters

For simplicity, we shall in the first approximation ignore the forecast estimation uncertainty and thus assume that the observed forecast errors $\hat{\mathbf{e}}_{t+1}$ do not differ from their optimal counterparts $\mathbf{e}^*_{t+1}$. To keep the notation simple, we hereafter denote the forecast errors by $\mathbf{e}_{t+1}$. The objective now is to estimate $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ from the sequences of observed forecast errors. Here, the sequence of times of change $t_1, \ldots, t_S$ is assumed to be known to the forecast evaluator.

Estimation is done following the ideas in (Andrews and Fair, 1988; Andrews, 1993). For this, we first need to generalize their setup, which allows only for a single shift (break). We start by splitting the sample of observed forecast errors into two parts,

$$D_1 \equiv [R, t_1 - 1] \cup [t_2, t_3 - 1] \cup \ldots \cup [t_S, T],$$
$$D_2 \equiv [t_1, t_2 - 1] \cup [t_3, t_4 - 1] \cup \ldots \cup [t_{S-1}, t_S - 1],$$

if S is even, with an analogous definition (that swaps the last two intervals) if S is odd. The idea is then to use a partial-sample GMM (PS-GMM) estimator that uses the data in D_1 to estimate $\boldsymbol{\tau}_1$ and the data in D_2 to estimate $\boldsymbol{\tau}_2$. To describe the estimator, suppose the true value of $\boldsymbol{\theta} \equiv (\boldsymbol{\tau}_1', \boldsymbol{\tau}_2')'$ is $\boldsymbol{\theta}_0 \equiv (\boldsymbol{\tau}_{10}', \boldsymbol{\tau}_{20}')'$. Note that the dimension of the parameter $\boldsymbol{\theta}$ equals $2n$ and that $\boldsymbol{\theta}$ takes values in $\Theta \equiv \mathcal{B}_q^n \times \mathcal{B}_q^n$. For the observations $t \in D_1$, we have the population orthogonality conditions

$$E[\mathbf{g}_p(\boldsymbol{\tau}_{10}; \mathbf{e}_{t+1}, \mathbf{x}_t)] = \mathbf{0}, \tag{12}$$

while for the observations $t \in D_2$ we have the population orthogonality conditions

$$E[\mathbf{g}_p(\boldsymbol{\tau}_{20}; \mathbf{e}_{t+1}, \mathbf{x}_t)] = \mathbf{0}, \tag{13}$$

where $\mathbf{x}_t$ is an $\mathcal{F}_t$-measurable d-vector of instruments, and $\mathbf{g}_p(\cdot; \mathbf{e}_{t+1}, \mathbf{x}_t) : \mathcal{B}_q^n \to \mathbb{R}^{nd}$ is the nd-vector-valued moment function given by

$$\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}, \mathbf{x}_t) \equiv \Big(p\boldsymbol{\nu}_p(\mathbf{e}_{t+1}) + \boldsymbol{\tau} \left\|\mathbf{e}_{t+1}\right\|_p^{p-1} + (p-1)\boldsymbol{\tau}'\mathbf{e}_{t+1} \left\|\mathbf{e}_{t+1}\right\|_p^{-1} \boldsymbol{\nu}_p(\mathbf{e}_{t+1})\Big) \otimes \mathbf{x}_t. \tag{14}$$

The idea is then to define an estimator

$$\hat{\boldsymbol{\theta}}_P \equiv (\hat{\boldsymbol{\tau}}'_{1P}, \hat{\boldsymbol{\tau}}'_{2P})'$$

of $\boldsymbol{\theta}_0 = (\boldsymbol{\tau}'_{10}, \boldsymbol{\tau}'_{20})'$ that is based on the sample analogues of these orthogonality conditions. Given p and π, $1 \leqslant p < \infty$ and $\pi \in (0,1)$, and given the observations $((\mathbf{x}'_R, \hat{\mathbf{e}}'_{R+1})', \ldots, (\mathbf{x}'_T, \hat{\mathbf{e}}'_{T+1})')'$, the PS-GMM estimator $\hat{\boldsymbol{\theta}}_P$ is defined as a solution to the minimization problem:

$$\min_{\boldsymbol{\theta}\in\Theta} \overline{\mathbf{g}}_p(\boldsymbol{\theta}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)' \mathbf{W} \overline{\mathbf{g}}_p(\boldsymbol{\theta}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t), \tag{15}$$

where $\overline{\mathbf{g}}_p(; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) : \Theta \to \mathbb{R}^{2nd}$ is a $2nd$-vector of stacked moment functions,

$$\overline{\mathbf{g}}_p(\boldsymbol{\theta}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) \equiv \frac{1}{P}\sum_{t\in D_1} \begin{pmatrix} \mathbf{g}_p(\boldsymbol{\tau_1}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) \\ \mathbf{0} \end{pmatrix} + \frac{1}{P}\sum_{t\in D_2} \begin{pmatrix} \mathbf{0} \\ \mathbf{g}_p(\boldsymbol{\tau_2}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) \end{pmatrix}, \tag{16}$$

$P = P_1 + P_2$ as before, and $\mathbf{W}$ is a positive definite symmetric $2nd \times 2nd$ weighting matrix. As the definition of the moment function $\overline{\mathbf{g}}_p$ in (16) indicates, the estimator $\hat{\boldsymbol{\theta}}_P$ consists of two components: $\hat{\boldsymbol{\tau}}_{1P}$, which uses the data in D_1 to estimate $\boldsymbol{\tau}_1$, and $\hat{\boldsymbol{\tau}}_{2P}$, which uses the data in D_2 to estimate $\boldsymbol{\tau}_2$. The estimator of Andrews and Fair (1988) is a special case of the above estimator obtained under a single shift (break).

We now establish the asymptotic distribution of the PS-GMM estimator $\hat{\boldsymbol{\theta}}_P$ for the case of no structural change. Let $\boldsymbol{\theta}_0 \equiv (\boldsymbol{\tau}'_0, \boldsymbol{\tau}'_0)'$ denote the true value of $\boldsymbol{\theta}$ in the case when no structural change occurs. The following quantities shall be of importance in the asymptotic variance of $\hat{\boldsymbol{\theta}}_P$:

$$\begin{aligned}
\mathbf{M}_{t+1} &\equiv \Big(p\boldsymbol{\nu}_p(\mathbf{e}_{t+1}) + \boldsymbol{\tau}\left\|\mathbf{e}_{t+1}\right\|_p^{p-1} + (p-1)\boldsymbol{\tau}'\mathbf{e}_{t+1}\left\|\mathbf{e}_{t+1}\right\|_p^{-1}\boldsymbol{\nu}_p(\mathbf{e}_{t+1})\Big) \in \mathbb{R}^n, \\
\mathbf{S} &\equiv E\left[\left(\mathbf{M}_{t+1}\mathbf{M}'_{t+1}\right) \otimes \left(\mathbf{x}_t\mathbf{x}'_t\right)\right] \in \mathbb{R}^{nd\times nd}, \\
\mathbf{B} &\equiv E\left[\left\|\mathbf{e}_{t+1}\right\|_p^{p-1}(\mathbf{I}_n \otimes \mathbf{x}_t) + (p-1)\left\|\mathbf{e}_{t+1}\right\|_p^{-1}(\boldsymbol{\nu}_p(\mathbf{e}_{t+1}) \otimes \mathbf{x}_t)\mathbf{e}'_{t+1}\right] \in \mathbb{R}^{nd\times n},
\end{aligned}$$

where $\pi = P_1/P$ as defined before. We shall consider the case where the weighting matrix $\mathbf{W}$ in (15) is optimally chosen as $\mathbf{W} = \mathbf{V}^{-1}$.

Theorem 1 *Let Assumption A1 and additional technical conditions hold. Then, given p, $1 \leqslant p <$*

∞, *and as* $(R, P_1, P_2) \to \infty$, *we have*

$$\sqrt{P}(\hat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}) \quad \textit{where} \quad \boldsymbol{\Omega} \equiv \begin{pmatrix} \frac{1}{\pi}(\mathbf{B}'\mathbf{S}^{-1}\mathbf{B})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{1-\pi}(\mathbf{B}'\mathbf{S}^{-1}\mathbf{B})^{-1} \end{pmatrix}.$$

B.3 Estimation of Cross-elasticities

We are now interested in measuring complementarities between different forecast errors under the multivariate loss $L(p, \boldsymbol{\tau}, \cdot)$ in 8. The idea is to measure the complementarities between say e_i and e_j by looking at the second cross derivative of the loss $L(p, \boldsymbol{\tau}, \cdot)$ in e_i and e_j. Once differentiating the loss $L(p, \boldsymbol{\tau}, \cdot)$ in Equation (8), we have:

$$\nabla_{\mathbf{e}} L(p, \boldsymbol{\tau}, \mathbf{e}) = p\boldsymbol{\nu}_p(\mathbf{e}) + \boldsymbol{\tau}\, \|\mathbf{e}\|_p^{p-1} + (p-1)\boldsymbol{\tau}'\mathbf{e}\frac{\boldsymbol{\nu}_p(\mathbf{e})}{\|\mathbf{e}\|_p}, \tag{17}$$

for all $\mathbf{e} \in \mathbb{R}^n$. Note that in the univariate case $n = 1$, the expression in Equation (17) reduces to $\nabla_e L(p, \tau, e) = [\tau + \mathrm{sgn}(e)]|e|^{p-1}$ (see Equation (8) in Elliott et al. (2005), p. 1121). Twice differentiating $L(p, \boldsymbol{\tau}, \cdot)$ then yields:

$$\begin{aligned} \Delta_{\mathbf{ee}} L(p, \boldsymbol{\tau}, \mathbf{e}) &= 2p\mathbf{V}_p(\mathbf{e}) + p(p-1)\mathbf{W}_p(\mathbf{e}) \\ &\quad + (p-1)\left[\frac{\boldsymbol{\tau}\boldsymbol{\nu}_p'(\mathbf{e}) + \boldsymbol{\nu}_p(\mathbf{e})\boldsymbol{\tau}'}{\|\mathbf{e}\|_p} + \frac{\boldsymbol{\tau}'\mathbf{e}}{\|\mathbf{e}\|_p}\left((p-1)\mathbf{W}_p(\mathbf{e}) - \frac{\boldsymbol{\nu}_p(\mathbf{e})\boldsymbol{\nu}_p'(\mathbf{e})}{\|\mathbf{e}\|_p^p}\right)\right], \end{aligned} \tag{18}$$

where we have used the fact that for any $1 \leqslant p < \infty$, $\dfrac{\boldsymbol{\tau}'\mathbf{e}}{\|\mathbf{e}\|_p}\mathbf{V}_p(\mathbf{e}) = \mathbf{0}$ for all $\mathbf{e} \in \mathbb{R}^n$. Note that in the univariate case $n = 1$, the Hessian in Equation (18) reduces to $\Delta_{ee} L(p, \tau, e) = 2\{p\delta(e)|e|^{p-1} + p(p-1)[1 + \tau\, \mathrm{sgn}(e)]|e|^{p-2}\}$ [see Equation (9) in Elliott et al. (2005) , p.1121].

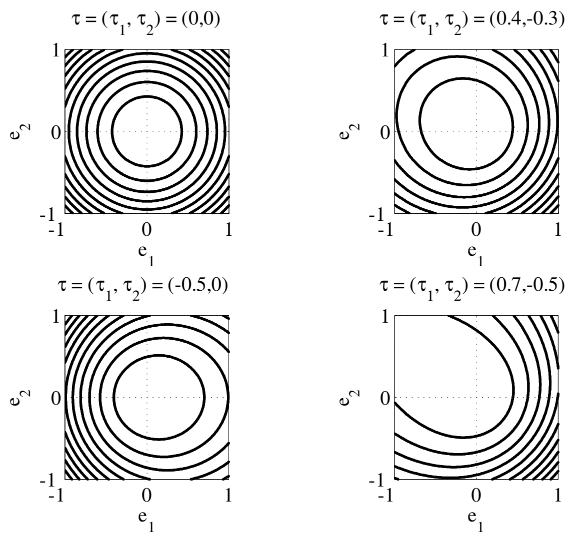
We are now interested in the off-diagonal elements of $E\left[\Delta_{\mathbf{ee}} L(p, \boldsymbol{\tau}, \mathbf{e})\right]$. For any pair $(i, j) : 1 \leq i < j \leq n$ let

$$\begin{aligned} \gamma_{ij} &\equiv E\left[\frac{\partial^2}{\partial e_i \partial e_j} L(p, \boldsymbol{\tau}, \mathbf{e})\right] \\ &= (p-1)E\left[\frac{\tau_i\, \mathrm{sgn}(e_j)|e_j|^{p-1} + \tau_j\, \mathrm{sgn}(e_i)|e_i|^{p-1}}{\|\mathbf{e}\|_p} - \boldsymbol{\tau}'\mathbf{e}\frac{\mathrm{sgn}(\mathrm{e_i})\,\mathrm{sgn}(\mathrm{e_j})\, |e_i|^{p-1}|e_j|^{p-1}}{\|\mathbf{e}\|_p^{p+1}}\right] \end{aligned} \tag{19}$$

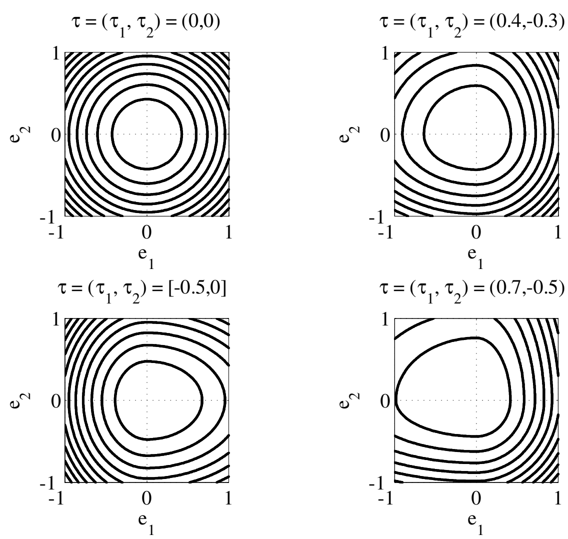
Note that γ_{ij} depends both on the values $p, \boldsymbol{\tau}$ of the multivariate loss function parameters, as well as on the forecast errors $\mathbf{e}$. It is worth pointing out that the magnitude of the entire vector of forecast errors determines the value of the complementarities in series i and j through the presence of terms $\|\mathbf{e}\|_p$ and $\boldsymbol{\tau}'\mathbf{e}$ in γ_{ij}.

For a given value of p, a consistent estimator $\hat{\gamma}_{ij}$ of γ_{ij} is simply obtained by plugging in the values of a consistent estimator $\hat{\boldsymbol{\tau}}$ of $\boldsymbol{\tau}$, and by replacing the expectations with their sample counterparts. In particular, the full sample estimator $\hat{\gamma}_{ij}$ of γ_{ij} is obtained as:

$$\hat{\gamma}_{ij} = (p-1)T^{-1}\sum_{t=R}^{T}\left[\frac{\hat{\tau}_i\,\mathrm{sgn}(\hat{e}_{jt})|\hat{e}_{jt}|^{p-1} + \hat{\tau}_j\,\mathrm{sgn}(\hat{e}_{it})|\hat{e}_{it}|^{p-1}}{\|\hat{\mathbf{e}}_{\mathbf{t}}\|_p} - \hat{\boldsymbol{\tau}}'\hat{\mathbf{e}}_{\mathbf{t}}\frac{\mathrm{sgn}(\hat{e}_{it})\,\mathrm{sgn}(\hat{e}_{jt})|\hat{e}_{it}|^{p-1}|\hat{e}_{jt}|^{p-1}}{\|\hat{\mathbf{e}}_{\mathbf{t}}\|_p^{p+1}}\right] \tag{20}$$

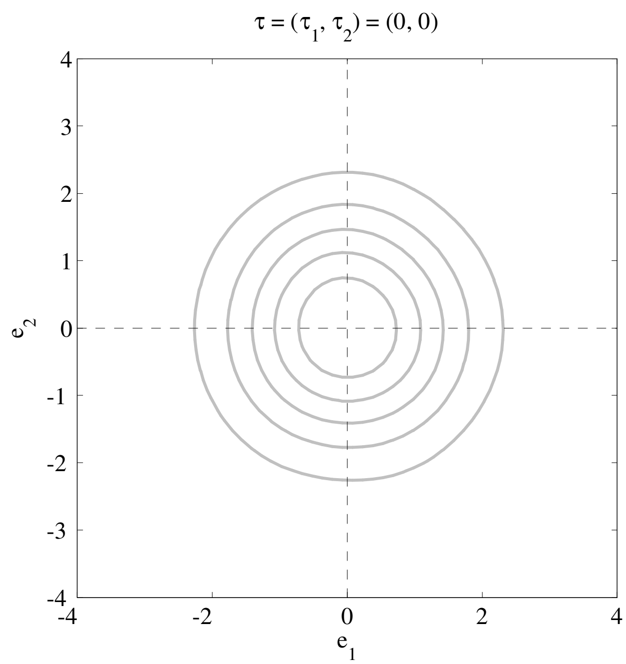


a) Separable Loss

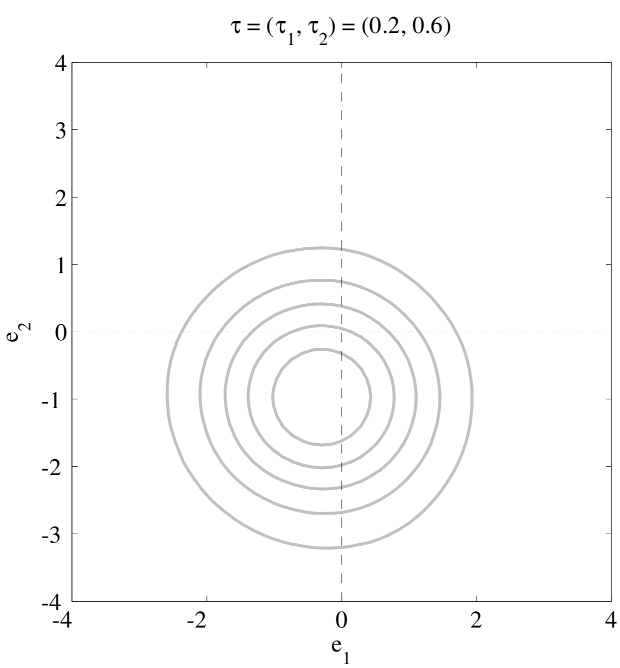


b) Non-separable Loss

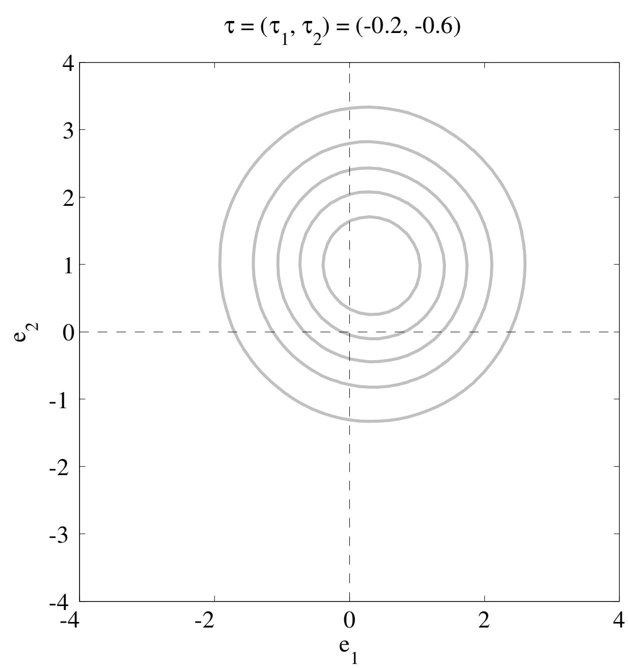
Figure 1. Iso Loss Contours. The figures show the iso-loss contours for various parameterizations of τ for the separable loss (eq. (2), top panel) and nonseperable loss (eq. (1), bottom panel). Each iso-loss contour represents combinations of forecast errors with equal loss.



(a) Symmetric loss.



(b) Higher losses from overestimation.



(c) Higher losses from underestimation.

Figure 2. Contour plots of the distribution of simulated forecast errors (kernel density smoothed). Shows the kernel density smoothed joint densities obtained from forecasts using the VAR (eq. 5) solving the problem in eq. 6.

	Nonseparable Loss			**Separable Loss**		
	Growth	**Inflation**	**UR**	**Growth**	**Inflation**	**UR**
tau	-0.29*	-0.03**	0.03***	-0.30*	-0.01*	0.32*
	(0.08)	(0.05)	(0.01)	(0.08)	(0.09)	(0.09)
J-stat		11.71*			10.89*	
p-value		0.07			0.09	

Table 1: Full sample estimates, $p = 2$.

Notes: The table shows the asymmetry parameters for the unemployment rate (UR), output growth, and inflation rate inferred using Greenbook forecasts for the period 1967:03-2011:12. Standard errors are shown in parentheses. p-values correspond to the 99^{th} percentile (***), 95^{th} percentile (**), and 90^{th} percentile (*). The J-stat tests the null of rationalizability of the forecasts and is detailed in Komunjer and Owyang (2012). p-values of the J-test correspond to a χ^2 distribution with 6 degrees of freedom.

	Nonseparable Loss		
	Volcker	**Great Moderation**	**Both**
Joint	54.75	7.84	14.97
	(0.00)	(0.05)	(0.00)
Growth	4.49	0.68	2.14
	(0.03)	(0.41)	(0.14)
Inflation	50.46	7.11	4.37
	(0.00)	(0.01)	(0.04)
UR	0.31	0.38	0.05
	(0.58)	(0.54)	(0.82)

	Separable Loss		
	Volcker	**Great Moderation**	**Both**
Joint	53.85	7.41	12.27
	(0.00)	(0.06)	(0.01)
Growth	5.19	1.07	1.3
	(0.02)	(0.30)	(0.25)
Inflation	51.78	6.68	3.84
	(0.00)	(0.01)	(0.05)
UR	0.52	0.6	8.17
	(0.47)	(0.44)	(0.00)

Table 2: Loss function subsample analysis: Wald tests.

Notes: The table contains the Wald statistic and the p-values (in parentheses) for the tests of breaks in the asymmetry parameters. The top panel contains tests using nonseparable loss; the bottom panel contains tests using separable loss. In each panel, the first row tests for a joint break in each variable's asymmetry parameter. The last three rows test for breaks in each of the three variables. For nonseperable loss, the test is for a break in the listed variable and no break in the other two variables. The first columns test for a single break at the Volcker appointment. The second columns test for a single break at the onset of the Great Moderation using only data from the post-Volcker period. The Volcker break is October 1979. The Great Moderation break is January 1984.

	Nonseparable Loss	
	Pre-Volcker	**Post- Great Moderation**
Growth	-0.11	-0.39
	0.13	0.09
Inflation	-0.29	0.07
	0.07	0.08
UR	0.03	0.04
	0.01	0.01
J-stat	5.70	5.91
	0.46	0.43

	Separable Loss	
	Pre-Volcker	**Post- Great Moderation**
Growth	-0.07	-0.40
	0.14	0.09
Inflation	-0.43	0.14
	0.10	0.13
UR	0.26	0.40
	0.11	0.12
J-stat	5.32	6.75
	0.50	0.35

Table 3: Subsample analysis: τ_n estimates, $p = 2$.

Notes: The table shows the asymmetry paraeters for the unemployment rate (UR), output growth, and inflation rate inferred using Greenbook forecasts for the pre-Volcker output growth, and inflation rate inferred using Greenbook forecasts for the pre-Volcker period (1967:03-1979:10, first column) and the Great Moderation period (1984:01-2011:12, second column). The asymmetry parameter is asymptotically normal. Standard errors are shown in parenthese. p-values correspond to the 99th percentile (***), 95th percentile (**), and 90th percentile (*). The J-stat tests the null of rationalizability of the forecasts and is detailed in Komunjer and Owyang (2012). p-values of the J-test correspond to a $/chi^2$ distribution with 6 degrees of freedom.

	growth-inflation	**growth-UR**	**inflation-UR**
Full Sample	-0.019	0.008	0.003
Pre-Volcker	-0.046	0.006	0.011
Post-GM	0.010	0.010	-0.002

Table 4: Complementarities across variables.

Notes: Complementarities are computed from estimates of the asymmetry parameters, as shown in equation (7).