



**ECONOMIC RESEARCH**  
FEDERAL RESERVE BANK OF ST. LOUIS  
WORKING PAPER SERIES

## How Persistent Are Unconventional Monetary Policy Effects?

<b>Authors</b>	Christopher J. Neely
<b>Working Paper Number</b>	2014-004F
<b>Revision Date</b>	April 2022
<b>Citable Link</b>	<a href="https://doi.org/10.20955/wp.2014.004">https://doi.org/10.20955/wp.2014.004</a>
<b>Suggested Citation</b>	Neely, C.J., 2022; How Persistent Are Unconventional Monetary Policy Effects?, Federal Reserve Bank of St. Louis Working Paper 2014-004. URL <a href="https://doi.org/10.20955/wp.2014.004">https://doi.org/10.20955/wp.2014.004</a>

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

The views expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment.

# How Persistent Are Unconventional Monetary Policy Effects?

Christopher J. Neely\*

This version: April 15, 2022

## Abstract

The weight of the evidence indicates that unconventional monetary policy (UMP) shocks had persistent effects on yields. To make this point, this paper illustrates that the most influential SVAR model of UMP effects, which implies transient effects, exhibits structural instability, sensitivity to specification and single observations that render the conclusions unreliable. Restricted SVAR models that limit asset return predictability are more stable and imply that UMP shocks were persistent. This conclusion is consistent with evidence from micro studies, surveys of professional forecasters, and quantity-of-debt models. Estimates of the dynamic effects of shocks should respect the limited predictability in asset prices.

*JEL classification: E430, E470, E520, C300*

*Keywords:* Federal Reserve; monetary policy; quantitative easing; large-scale asset purchase; VAR; forecasting; structural breaks; good deal

---

\*Corresponding author. Send correspondence to Chris Neely, Box 442, Federal Reserve Bank of St. Louis, St. Louis, MO 63166-0442; e-mail: [neely@stls.frb.org](mailto:neely@stls.frb.org); phone: +1-314-444-8568; fax: +1-314-444-8731. Christopher J. Neely is a Vice President and economist at the Federal Reserve Bank of St. Louis. I thank Michael Bauer, Graham Elliot, John Keating, Clemens Kool, Leo Krippner, Mike McCracken, Valerio Potì, Gene Savin, Carsten Trenkler, Paul Wilson, Mark Wohar, participants at the Missouri Economic Conference, the FRB St. Louis brown bag meetings, University of Iowa, Federal Reserve System Committee on Macroeconomics, Midwest Macro Meetings, the EFMA 2014 meetings, and the NBER-NSF Time Series conference for helpful discussions and Brett Fawley, Sean Grover, Michael Varley, Evans Karson, Jacob Haas, and Sam Jordan-Wood for research assistance. The views expressed in this paper are those of the author and do not necessarily reflect those of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks.

The financial market turmoil that followed Lehman Brothers' September 2008 bankruptcy prompted extraordinary measures from monetary authorities. The Federal Reserve engaged in emergency lending and reduced the federal funds rate to very low levels by December 16, 2008. These measures failed to stem the economic slide, however, and the Federal Reserve soon turned to forward guidance and broad asset purchases to support the economy, especially housing markets. Quantitative easing (QE) purchases occurred in phases: QE1 was announced on November 25, 2008, and March 18, 2009; QE2 on November 3, 2010, a maturity extension program ("Operation Twist") on September 21, 2011; and QE3 on September 13, 2012. Through these programs, the Fed purchased trillions of dollars of long-term assets.

Event studies have established that QE announcements strongly affected asset prices. Gagnon et al.'s (2011) find that the Fed's 2008-09 QE announcements reduced U.S. long-term yields. Joyce et al. (2011) find quantitatively similar effects for U.K. QE. Neely (2015) evaluates QE1's effect on international long yields and exchange rates. Bauer and Neely (2014) evaluate the relative importance of signaling and portfolio balance channels on international interest rates with term structure models. Bhattarai and Neely (forthcoming) survey the UMP literature.

There has been much less work on the impact of QE on macroeconomic variables (Baumeister and Benati, 2013; Gambacorta, Hofmann, and Peersman, 2014; and Gertler and Karadi, 2013). A significant difficulty with such research is that the macro effects of QE depend on the persistence of the asset price effects of QE. Transient QE shocks to interest rates presumably imply that QE is a much less effective policy than would persistent QE effects.

It is difficult, however, to estimate the persistence of unconventional monetary policy (UMP) shocks because it implicitly requires accurately estimating a counterfactual path—a path in the absence of the policy shock—for asset prices. Wright (2012) suggests measuring the persistence

of monetary shocks with a structural vector autoregression (SVAR) estimated on 6 daily U.S. yields and inflation compensation series, using heteroskedasticity to identify monetary shocks (Rigobon and Sack, 2004).<sup>1</sup> The SVAR implies that UMP shocks have large, but very transient, effects on U.S. interest rates; most of the impact dissipates within 6 months. This is consistent with the views of many market observers who concluded that QE failed because long yields rose as the Fed was expanding (Greenlaw et al., 2018).

The academic literature comes to no clear consensus on persistence (Kuttner, 2018). Many influential works have cited Wright's (2012) estimates of UMP shock transience, without necessarily endorsing those estimates: Bernanke (2012), Joyce et al. (2012), Gilchrist and Zakrajšek (2013), and Hansen and McMahon (2016). Using other methods and data, Belke, Gros, and Osowski (2017) and Gabriel and Lutz (2017) found transient effects, while Swanson (2021), Altavilla and Giannone (2017) and Altavilla et al. (2019) use factor models on different data sets to produce evidence of persistent UMP shocks.

Ubiquitous overfitting and structural instability in asset pricing models provide good a priori reasons to question estimated dynamic patterns, i.e., transient shocks, in asset prices. Meese and Rogoff (1983) showed that structural exchange rate models produced very poor out-of-sample (OOS) performance. Neely and Weller (2000) show that inferring long-run asset price behavior from VARs is unreliable (Bekaert and Hodrick, 1992). Faust, Rogers, and Wright (2003) convincingly argue that Mark's (1995) foreign exchange forecasting model is fragile with respect to data vintages. Welch and Goyal (2008) question the usefulness of traditional equity premium predictors in OOS forecasting exercises.

Informed by these lessons, the present paper makes two related points: a conclusion about

---

<sup>1</sup> Mamaysky (2018) looked at shorter-horizon, dynamic equity market reactions to UMP announcements.

UMP-shock persistence and a caution about empirical methods and theoretical restrictions. First, the most influential model on the topic, i.e., Wright (2012), does not provide reliable evidence that UMP shocks are transient. Quantity-of-debt analyses, micro studies of portfolio choice, and surveys of professional forecasters indicate that UMP shocks are persistent. Second, empirical models should be structurally stable and not overly sensitive to the removal of a single observation. Excessive asset price predictability is a red flag as it could reflect problems such as instability from overfitting or sensitivity. Rational asset pricing and reasonable risk aversion imply that shocks have fairly persistent impacts on bond yields and asset prices.<sup>2</sup>

We start by critically examining the most influential paper on UMP shock persistence, that of Wright (2012), to illustrate problems that render the conclusions unreliable. Specifically, the first-order structural vector autoregression (SVAR(1)) is structurally unstable and forecasts poorly, indicating misspecification. Minor changes to the specification —i.e., lag length, removal of an observation—substantially alter point estimates of shock persistence. This paper also argues that rational asset pricing and reasonable risk aversion rule out the unexploited, risk-adjusted expected returns that would accompany transient effects. Models that are consistent with reasonable risk aversion and rational asset pricing forecast better than the SVAR and imply more persistent responses to monetary shocks.

In critiquing the SVAR(1), we compare it to alternative models that are not necessarily consistent with each other. For example, SVARs with longer lag lengths, e.g., the SVAR(5), is not consistent with a martingale model. The inconsistency is not a problem because the goal of the present paper is not to find a specific model to estimate the shock persistence. Rather, the SVAR(1)'s deficiencies exemplify problems inherent in such exercises. While we seek the best

---

<sup>2</sup> Although this paper does not specifically address the estimation of full macro models, it seems that theoretical restrictions, such as limited asset price predictability, should inform those methods as well as other empirical work.

way to estimate empirical relations, we recognize the limits of data and methods. Atheoretic time series models leave much uncertainty about the persistence of monetary shocks, particularly using short samples.

Finally, this paper briefly reviews the evidence from other sorts of models, such as regression models that explain yields or term premia with quantity-of-debt held by the public, micro-studies of lending or portfolio choice, and surveys of professional forecasters. This evidence supports the view that UMP shocks have persistent effects on long yields.

The next section of the paper describes Wright's SVAR methodology. Section 3 replicates Wright's main findings, while Section 4 illustrates structural instability and sensitivity problems. Section 5 shows that the baseline VAR is inconsistent with rational asset pricing and reasonable risk aversion, while Section 6 presents restricted VARs that are consistent with rational pricing and reasonable risk aversion. Section 7 discusses alternative evidence for the persistence of UMP shocks, and Section 8 concludes.

## 2. The Structural VAR Methodology

The reduced-form VAR can be written as follows:

$$A(L)y_t = \varepsilon_t, \quad (1)$$

where  $A(L)$  is a polynomial in the lag operator,  $y_t$  is the vector of endogenous variables, and  $\varepsilon_t$  denotes the reduced-form error vector, which is related to the structural errors:  $\varepsilon_t = \sum_{i=1}^6 R_i u_{t,i}$ , where  $R_i$  is a  $6 \times 1$  vector of the initial impacts of the  $i$ th structural shock,  $u_{t,i}$ , on each of the  $y_t$  elements. The reduced-form covariance matrix is a function of structural parameters:

$$\Sigma = \sum_{i=1}^6 R_i R_i' o_i^2, \quad (2)$$

where  $o_i^2$  denotes the variance of the  $i$ th structural shock. The moving average representation for

the  $i$ th structural shock would be  $(I - A(L))^{-1}R_i$ .<sup>3</sup>

Ordinary least squares (OLS) estimates of coefficients on lagged endogenous regressors will be biased in finite samples and will generally underestimate the data's persistence. Therefore, Wright follows Kilian (1998) in correcting this bias with a bootstrapping procedure.<sup>4</sup> Appendix A describes the bias correction and bootstrapping methods used in this paper.

Wright (2012) follows the spirit of Rigobon and Sack's (2004) identification-through-heteroskedasticity procedures that the latter use to estimate the effect of monetary policy shocks on asset prices. Wright assumes only that the variance of the structural monetary policy shock,  $u_{t,1}$ , is higher on 28 specific monetary announcement days ( $\sigma_{1,A}^2$ ) than on non-announcement days ( $\sigma_{1,N}^2$ ), which creates heteroskedastic reduced-form errors,  $\varepsilon_t$ .<sup>5</sup> The announcement set included dates of Federal Open Market Committee (FOMC) meetings and other announcements or speeches by the Chairman that were relevant to UMP.<sup>6</sup> Under this assumption and using (2), the difference in the residual reduced-form covariance matrices on announcement and non-announcement days is a function of the initial impact vector,  $R_1$ , of monetary shocks:

$$\Sigma_1 - \Sigma_0 = R_1 R_1' (\sigma_{1,A}^2 - \sigma_{1,N}^2). \quad (3)$$

---

<sup>3</sup> A VAR is atheoretic. Duffee (2013) surveys the literature on forecasting interest rates with dynamic models that impose economic theory in the form of no-arbitrage conditions. He argues that more economic theory is needed to pin down risk premia dynamics.

<sup>4</sup> Several types of bootstrapping, all of which incorporated the heteroskedastic structure of the data-generating process, provided fairly similar results. The results presented here use a bootstrap drawn from two distributions, a 10-day moving block bootstrap for non-monetary policy days, overlaid with draws on the monetary policy days from the distribution of residuals on those days.

<sup>5</sup> The normalization that the monetary policy shock is the first structural shock is innocuous and does not affect any results. It is not related to the ordering of variables in a VAR under a Cholesky factorization.

<sup>6</sup> The monetary announcement dates were as follows: 11/25/2008, 12/1/2008, 12/16/2008, 1/28/2009, 3/18/2009, 4/29/2009, 6/24/2009, 8/12/2009, 9/23/2009, 11/4/2009, 12/16/2009, 1/27/2010, 3/16/2010, 4/28/2010, 6/23/2010, 8/10/2010, 8/27/2010, 9/21/2010, 10/15/2010, 11/3/2010, 12/14/2010, 1/26/2011, 3/15/2011, 4/27/2011, 6/22/2011, 8/9/2011, 8/26/2011, and 9/21/2011. Wright's Table 5 incorrectly lists June 2, 2011, as a monetary policy announcement date. The correct date should be June 22, 2011, which was the date of an FOMC meeting.

Estimates of  $A(L)$ ,  $\Sigma_1$ , and  $\Sigma_0$  enable one to estimate  $R_1$  from (3). Because the terms in the product  $R_1 R_1' (o_{1,A}^2 - o_{1,N}^2)$  are not separately identified, Wright normalizes  $(o_{1,A}^2 - o_{1,N}^2)$  to 1 and solves for the elements of  $R_1$  by minimizing the quadratic function of the difference vector,  $vech[\hat{R}_1 \hat{R}_1' - (\hat{\Sigma}_1 - \hat{\Sigma}_0)]$ , using the covariance matrix of  $(\hat{\Sigma}_1 - \hat{\Sigma}_0)$  to appropriately weight the moments. Estimates of  $A(L)$  and  $R_1$  permit one to construct impulse responses for the UMP shocks. There is no need for additional identifying assumptions to investigate UMP shocks.

Wright (2012) block bootstraps the VAR system to test two hypotheses: 1) The covariance matrices are the same on announcement and non-announcement days (i.e.,  $\Sigma_1 = \Sigma_0$ ), and 2) there is a single monetary policy shock (i.e.,  $R_1 R_1' = (\Sigma_1 - \Sigma_0)$ ). The bootstrapping tests, which are implicitly conducted under the assumption of a VAR system whose first two moments are stable, reject the null that  $\Sigma_1 = \Sigma_0$  but fail to reject that there is a single monetary policy shock.

### 3. Data and Replication of SVAR Results

#### *3.1 Replication of Wright (2012)*

Wright (2012) estimates a 1-lag VAR, using the bias-adjusted bootstrap of Kilian (1998), with 6 daily U.S. interest rates—the 2- and 10-year nominal Treasury yield; the 5-year and 5-to-10 year forward inflation compensation; and the Moody’s Baa- and Aaa-rated corporate bond yield indices—using daily data from November 3, 2008, to September 30, 2011. This paper replicates Wright’s results with similar data, procedures, and identification scheme.

The reduced-form VAR coefficients and the initial impact of the structural shocks determine the impulse response functions. The moving average representation can be written as follows:

$$y_t = A(L)^{-1} \varepsilon_t = (I - A_1 L)^{-1} R u_t, \quad (4)$$

where  $A_1$  is the matrix of reduced-form VAR coefficients and  $R$  is the  $6 \times 6$  matrix relating the structural error vector,  $u_t$ , to the reduced-form error vector,  $\varepsilon_t$ .  $R$ ’s first column is  $R_1$ . In



calculating the impulse responses, Wright normalizes the monetary shock to reduce 10-year yields by 25 basis points on impact.

Figure 1 illustrates the impulse responses and 90 percent bootstrapped confidence intervals, which are similar to — perhaps a bit wider than — those in Wright’s paper. UMP shocks significantly change 10-year Treasury, Baa, and Aaa rates, with immediate effects on the corporate rates ranging from 40 to more than 100 percent of Treasury changes.<sup>7</sup> This immediate effect is consistent with event studies of UMP (e.g., Gagnon et al., 2011 and Neely, 2015).

The significant initial effects of the monetary policy shock wear off very quickly, however. The point estimates of the half-lives of the 10-year Treasury and corporate yields responses range from 2 to 6 months in Figure 1. Although one cannot reject that the half-lives are at least a year, Wright focuses on the point estimates: “[T]he impulse responses on 10-year Treasuries and corporate yields are statistically significant but only for a short time. The half-life of the estimated impulse responses for Treasury and corporate yields is two or three months.” — Wright (2012, page F452). Readers have understandably followed this interpretation of the results, e.g., Joyce et al. (2012), Gagnon (2016), and Yu (2016), which suggests that UMP actions have only very transient effects on yields and therefore very modest effects on macroeconomic variables. Wright (2012, p. F465) summarizes: “*To the extent that longer term interest rates are important for aggregate demand, unconventional monetary policy at the zero bound has had a stimulative effect on the economy but it might have been quite modest.*”

### ***3.2 The effect of model uncertainty***

The confidence intervals in Figure 1 imply fairly precise estimates of dynamic behavior. But

---

<sup>7</sup> The bias-corrected coefficients that produced Figure 1 do imply a stationary system, but the response point estimates are sensitive to small changes in the coefficients near the stationarity boundary, especially at long horizons. Therefore, the point estimates for the impulse responses can easily fail to coincide with the median of the distribution of responses.

these confidence intervals are both pointwise—that is, narrower than uniform confidence bands—and conditional on the VAR lag length. Such conditional confidence intervals disguise any uncertainty about the true model/lag length, suggesting a misleading degree of precision. It is therefore worth considering the effect of model (i.e., lag length) uncertainty on inference.

Wright (2012) chose a VAR lag length of 1 to minimize the Bayesian information criterion (BIC) that has a strong preference for parsimony. Intuitively, a VAR(1) seems unlikely to accurately characterize the impact of shocks at long horizons, as the time path of the estimated responses will be a function of just a few first-order covariances. A 3-year sample will have very little information about long-run relationships, and so larger models that might govern those relationships will be estimated imprecisely and discarded by the BIC.

Even empirically, however, a VAR(1) appears to be insufficient. Ljung-Box Q tests on the residuals from 1- and 2-lag VARs often reject the null of no autocorrelation, suggesting that VARs with these lag lengths are misspecified and more than 2 lags are needed. Indeed, the Akaike information criterion prefers a lag length of 3, although 2, 3, 4, and 5 lags outperform 1 lag by this criterion.<sup>8</sup> Full results are omitted for brevity but are available on request.

This model selection exercise points out a decision theoretic problem with econometric practice. Applied econometricians typically search for a parsimonious model to avoid overfitting. Insignificant parameters are generally assumed to be economically unimportant. But statistical tests can only reject or fail to reject null hypotheses, they cannot “accept” them. Failing to reject the null does not imply that it is correct (see Cochrane (2007)). The assumption that imprecisely

---

<sup>8</sup> The reader might think that a 1-lag structure would be most consistent with term structure models, which are often Markovian. Markovian term structure models take that structure for reasons of tractability/parsimony, rather than consistency with economic theory. The well-known Heath-Jarrow-Morton term structure framework allows for non-Markovian dynamics under the physical measure, even while the risk-adjusted dynamics remain Markovian. It is well known that additional lags (Cochrane and Piazzesi, 2005; Joslin, Le, and Singleton, 2013) or moving average terms (Feunou and Fontaine, 2018) can improve the dynamic fit of term structure models.

estimated parameters from a short sample are exactly zero can affect economic inference, particularly for sensitive, nonlinear functions such as impulse responses.

Because economic inference can be very sensitive to lag length, it is worth examining impulse responses generated by VARs with longer lag lengths. The top panel of Figure 2 shows the point estimates of the impulse responses for VARs estimated with lag lengths of 1, 3, 5, and 10 lags. This top panel omits confidence intervals to focus on the pattern in persistence by lag. The graph shows that persistence monotonically rises with VAR lag length for the 10-year yield. For the 5- and 10-lag models, the increase in point estimate persistence is very substantial.<sup>9</sup> One obvious interpretation of these results is that, if the coefficients on higher lags are truly “small” compared to the precision with which they can be estimated, then the BIC will set them to zero. This does not mean that these coefficients are actually zero, merely that their contribution to 1-step ahead forecasting in the 3-year sample is too modest for the BIC. These “small” coefficients greatly increase the estimated persistence in the VAR, however.

The dependence of the economic inference on lag length raises the concern that the BIC, which values parsimony, might choose an incorrect model. To determine the likelihood of incorrectly choosing a 1-lag model using the VAR yield data, we simulated 1000 data sets from VARs with higher lag orders—2, 3, 5 and 10 lags—using pseudo-true VAR coefficients that were estimated from the real data. We then compared the BIC for 1-lag and N-lag VAR models on the simulated data sets. The BIC incorrectly chose the 1-lag model over the correct N-lag model a very high proportion of times: 95, 55, 91 and 94 percent of the time for 2-, 3-, 5- and 10-lag models, respectively. Thus—conditional on a higher order VAR—the model selection procedure is very likely to distort inference toward choosing a lower lag length and inferring

---

<sup>9</sup> Although the VAR(10) impulse responses appear to show potential nonstationary behavior, examination of very long-horizon behavior confirms that the system is stationary.

transient shocks. Full results are available from the author.

This exercise does not imply that the BIC is a bad model selection tool. With an infinite amount of data from a stable data generating process, the BIC will pick the correct model. Rather, this exercise shows that with only a relatively short sample of noisy data, the BIC — which fits 1-step ahead forecasts — tends to pick small models that will not necessarily describe the long-run dynamics well but will favor transient shocks.

One can partially account for such uncertainty within a finite model set by model averaging over VARs of various lengths, weighting each model's parameters by the model's BIC, as suggested by Buckland, Burnham, and Augustin (1997).<sup>10</sup> The top panel of Figure 2 illustrates the point estimate of the impulse response of the 10-year Treasury from the average estimate of a set of VARs from 1 to 15 lags. The bottom panel of Figure 2 displays the point estimates and 90 percent confidence intervals from the 1-lag and Buckland-Burnham-Augustin-averaged model. The averaged model implies a much more persistent response with wider confidence intervals than does the 1-lag model. The half-life of the shock more than doubles and one cannot reject the hypothesis of no diminution in the shock for more than a year.<sup>11</sup> In other words, formally accounting for model uncertainty substantially increases the estimated persistence of the UMP shocks. The next sections of the paper, however, argue that the VAR(1) exhibits more serious problems that imply we cannot draw conclusions on persistence from this VAR.

## 4. Analysis of the VAR's Stability

The first two moments of the estimated VAR must be stable over time or the impulse responses in Figure 1 are spurious and the inference of transient effects is unreliable. Although

---

<sup>10</sup> Wang, Zhang, and Zou (2009) usefully review the literature on frequentist model averaging.

<sup>11</sup> Note that the relative positions of the point estimate and the 5 percent path shows that the distribution is skewed left, reminiscent of the well-known left skew in sampling distributions of persistence in univariate autoregressive processes.

unrestricted VARs are not necessarily the best forecasting models, they must describe stable dynamic relations to accurately characterize the responses of variables to shocks. But VARs and other time-series relations are notoriously unstable predictors of asset prices (Rossi, 2013, and Stock and Watson, 2003). Empirical models have failed to forecast sundry asset prices in OOS exercises: exchange rates (Meese and Rogoff, 1983; Faust, Rogers, and Wright, 2003); equities (Welch and Goyal, 2008); interest rates (Thornton and Valente, 2012), and cross-asset studies (Neely and Weller, 2000).

#### ***4.1 Forecasting exercises***

Econometric tests, as in Andrews (1993) or Bai and Perron (1998, 2003), constitute the most powerful tests for structural stability, but OOS forecasting exercises provide an informal and intuitively attractive supplement to formal tests (Rapach and Wohar, 2006). Therefore, before formally testing the stability of the VAR, this paper first considers whether the VAR forecasts outperform a no-change (i.e., martingale) benchmark in OOS testing.

To investigate the OOS forecast performance of the VARs, we estimate the coefficients with in-sample data (2008:11:03–2011:09:30) to forecast each of the variables in the system over the OOS period (2011:10:01–2013:11:27) at horizons of 1, 20, 60, and 120 days. At each date in the OOS period, we condition on the actual data at date  $t$  and the parameters as estimated over the fixed sample period and project the path of the system at dates  $t + 1$  through  $t + 120$ . We then update the data for the next period's set of forecasts. This provides a set of 538 one-period-ahead forecasts, 519 overlapping 20-period-ahead forecasts, 479 overlapping 60-period forecasts, and 419 overlapping 120-period forecasts.<sup>12</sup>

Table 1 shows the OOS root mean square forecasting errors (RMSFEs) in basis points, over

---

<sup>12</sup> The overlapping  $n$ -period forecast errors will have at least an  $n-1$ -order serial correlation that the tests must account for.

1-, 20-, 60- and 120-day horizons, for a naïve, no-change model for the interest rates and the bias-adjusted VAR, respectively. The third panel of Table 1 shows the ratio of those RMSFEs, the Theil U-statistics. Theil ratios less than 1 favor the VAR model; ratios greater than 1 favor the naïve model. The bottom panel of Table 1 shows the proportion of the bootstrapped Theil statistics that exceed the real Theil statistics under the null that the VAR generated the data.

The VAR's OOS forecast performance is poor. A naïve, no-change prediction is superior to the VAR forecasts for 19 of 24 horizon-variable combinations considered (panel 3, Table 1). The naïve, no-change forecast outperforms the VAR at every horizon for every yield variable.<sup>13</sup> This strongly suggests that the VAR does not accurately model dynamic relations, and the impulse response functions in Figure 1 do not reliably describe dynamic relations.<sup>14</sup>

It is true that misspecified models can forecast better than correctly specified models, out-of-sample. Generally, however, this is only true when the correctly specified model's parameters are estimated so poorly in a finite sample that the misspecified model describes the dynamic relations better than the correctly specified, but poorly estimated, model. A correctly specified linear model with the true parameters will outforecast a misspecified model. In the present case, the naïve model clearly outperforms the VAR(1) in every yield equation, at every horizon, indicating that the estimated VAR(1) describes the dynamics very poorly.

#### ***4.2 Formal structural stability tests***

Structural instability, a form of model misspecification, is common in time-series regressions. Formal econometric tests are more powerful tests of stability than OOS forecasting

---

<sup>13</sup> Tightening up the priors on a Bayesian VAR could reproduce the naïve, martingale forecasts (Litterman, 1986), but then one would obtain very persistent impulse responses, not the mean-reverting impulse responses that indicate transient effects.

<sup>14</sup> Although we omit full results for brevity, we also examined forecast bias. The naïve predictions are never systematically biased in a statistically significant way, but the VAR yield predictions are biased at all horizons.

exercises. To test for structural instability, we follow Andrews (1993) by calculating the Wald test statistics for a structural break in the uncorrected VAR coefficients at each observation in the middle third of each sample.<sup>15</sup> Newey-West covariance matrices are calculated with automatic lag length selection (Newey and West, 1994). The supremum of these test statistics identifies a possible structural break in the series but will have a nonstandard distribution (Andrews, 1993). The critical values for the supremum are calculated from a Monte Carlo simulation using a moving block bootstrap with a window of length 10.

Consistent with this poor OOS forecasting performance, the top panel of Figure 3 plots the Andrews (1993) unknown-point structural break statistics for the null that the VAR parameters are stable over time, along with 1, 5, and 10 percent critical values. The structural break statistics are often well above the 5 percent critical value—particularly during the QE1 period—rejecting the null of stable VAR parameters. The intertemporal instability of the VAR indicates that the VAR impulse response functions in Figure 1 are spurious and the inference from them is suspect.

To determine the prevalence of breaks in the six individual VAR equations during the sample period, one can conduct Bai-Perron (1998, 2003) tests for breaks at unknown points for the uncorrected VAR estimates.<sup>16</sup> Bai and Perron (2003) recommend that one first test for the presence of any breaks with the *UD max* or *WD max* tests and then evaluate the number of breaks by sequentially testing up for the maximum number of breaks with the SupF tests.<sup>17</sup>

This paper follows those Bai-Perron guidelines under the assumptions of a maximum of 3

---

<sup>15</sup> We construct the Andrews test statistics for the uncorrected VAR coefficients to keep computational cost within reasonable bounds. The bias correction is very unlikely to change the outcome of the structural stability tests.

<sup>16</sup> Because the Andrews (1993) and Bai and Perron (1998, 2003) structural stability tests do not require one to specify the date of the break, they generally have much less power to reject the null of stability than tests that do so. Tests for a break-at-an-unknown-date typically require large separation between the models, i.e., big breaks.

<sup>17</sup> Denoting the maximum number of breaks permitted by  $M$ , the *UD max* statistic tests for a break by considering whether the maximum of all  $M$   $F$ -statistics exceeds its critical value, while the *WD max* statistic also tests for breaks with a weighted average of the  $F$ -statistics in which the marginal  $p$ -values are equalized across statistics. See Bai and Perron (1998, 2003).

breaks with at least 20 percent of the original sample between each break. The first and second rows of Table 2 show that the *UD max* and *WD max* statistics reject the null of no breaks for all equations at conventional significance levels. The third and fourth rows of Table 2 illustrates that one can reject the null of 1 break in favor of 2 breaks for 4 of the 6 equations at the 5 percent level (row 4), but one cannot reject the null of 2 breaks in favor of 3 breaks (row 3). In summary, all equations exhibit at least one break and most exhibit at least two breaks.

The estimates of the VAR regression coefficients,  $A_1$ , drive the poor performance on structural stability tests and OOS forecasting in Tables 1 and 2 and Figure 3. The identification-through-heteroskedasticity of the mapping between reduced form and structural errors, i.e.,  $R_i$ , does not affect these exercises. Although another identification scheme for the contemporaneous impact matrix would alter the estimated impulse response functions, no conventional identification of that contemporaneous matrix will make those response functions reliable because they will still depend crucially on the unstable reduced form regression coefficients in the VAR. Therefore, no alternative identification for the contemporaneous impacts will create reliable impulse responses.

#### ***4.3 Does the yield data or the sample period create instability?***

The highly significant break statistics in the top panel of Figure 3 raises the question of whether it is the nature of the yield data or the particular sample that generated such instability. To investigate this question, one can compare the break statistics for the yields during 2008-11 with those from a VAR on the same data during a calmer sample, 1999-2006, and on dissimilar data—monthly macro data—from 1983 to 2006. These samples were chosen to coincide with the “Great Moderation.” The macro data have been commonly used in VAR studies and include



industrial production (IP), the consumer price index for all urban consumers (CPI-U [CPI]), personal consumption expenditures (PCE), the 3-month Treasury yield (3M), the price of West Texas Intermediate crude (WTI), and the civilian unemployment rate (UR).

Bias-corrected VAR(1) models were estimated on both datasets, and Andrews break statistics and critical values were constructed with Newey-West covariance matrices with automatic lag length selection and a 10-day moving block bootstrap to simulate data. The center and bottom panels of Figure 3 display the break statistics for the VARs. Neither system shows clear evidence of instability, though the macro break statistics do approach the 10 percent region near the end of the sample. This suggests that a VAR on yields is not necessarily unstable but that the turbulent conditions during the 2008:11–2011:09 sample were likely an important factor in the instability.

#### ***4.4 Sensitivity analysis***

The identification strategy of the SVAR depends on the highly kurtotic sample of just 28 shocks around monetary policy announcements, some of which are very large. The shock of March 18, 2009 is almost 4 standard deviations.<sup>18</sup> To investigate the robustness of the persistence results, we reclassified March 18, 2009 as a non-announcement day and reestimated the SVAR. This specification implies much more persistence in effects on yields and much wider confidence intervals (Figure 4).<sup>19</sup> (Note that the vertical axes in Figure 4 are twice those of Figure 1). The estimated half-life of a UMP shock to the 10-year Treasury yield exceeds a year in Figure 4.

This subsection showed that the VAR that produced Figure 1—evidence for the transient

---

<sup>18</sup> Swanson (2021) notes that the exclusion of this March 18 observation from his factor-model affects the estimated persistence of unconventional asset purchases.

<sup>19</sup> The point estimates for the impulse response functions in Figure 4 deviate somewhat from the medians of the confidence intervals because they are both estimated as elements of  $R_1$ , which are chosen to minimize an overidentified relation involving the covariance matrices of VAR shocks: the weighted quadratic function of the difference vector,  $vech[\hat{R}_1 \hat{R}_1' - (\hat{\Sigma}_1 - \hat{\Sigma}_0)]$ . This estimation is very sensitive to the announcement list.

effects of UMP shocks—forecasts very poorly OOS, is structurally unstable, and is sensitive to the exclusion of one observation. The relative forecasting success of the martingale model indicates that persistent shocks better approximate the dynamic structure. The next section investigates the relation of rational asset pricing to shock persistence.

## 5. Is the Estimated Predictability Consistent with Rational Pricing?

In a world of rational, unconstrained risk-neutral investors, expected excess returns should be bid to zero and systematically transient shocks are impossible because they would create unexploited profit opportunities that would be inconsistent with the efficient markets hypothesis.<sup>20</sup> But non-zero expected excess returns—including systematically transient shocks—are potentially consistent with risk-averse investors and greater risk aversion permits more potential predictability. This section asks if the predictability in the VAR—i.e., mean reverting impulse responses—is consistent with rational pricing and reasonable risk aversion.<sup>21</sup>

Potì and Siddique (2013) show that the product of the square of the coefficient of risk aversion and the variance of the market return must exceed the  $R^2$  from a predictive regression.<sup>22</sup>

$$R^2 \leq (1 + R_f)^2 RRA_V^2 \sigma^2(r_{m,t+1}) \cong RRA_V^2 \sigma^2(r_{m,t+1}), \quad (5)$$

where  $R_f$  is the riskless rate,  $RRA_V$  is the upper bound on relative risk aversion (RRA), and

---

<sup>20</sup> Bernanke (2020) argues that transient price shocks are not intuitively persuasive because smart investors could arbitrage away such patterns betting on mean reversion. Duffie (2010) makes a careful case that immobile capital limits arbitrage and might produce some asset price dynamics. Sporadic trading in thin markets might produce expected returns over a horizon of a few days. Longer effects might come from systematic overreactions to certain types of news. In extreme cases with significant limits to arbitrage—e.g., the case of an insurance company whose capital was depleted by disaster—adjustment might take years because participants rationally delay replenishing capital. It is not clear how such examples would explain transient UMP effects in closely watched, thickly traded, international bond markets, however.

<sup>21</sup> The exercise in this section could be interpreted as applying Manski’s (1993, 2010) method of partially identifying theory-driven restrictions to macroeconomics / microfinance.

<sup>22</sup> Kirby (1998) first formalized the intuition that the willingness to substitute consumption across states of nature must bound the  $R^2$  from a predictive regression of asset returns. Appendix B summarizes the implications of Potì and Siddique’s (2013) work for the present paper.

$\sigma^2(r_{m,t+1})$  is the variance of the market excess return,  $r_{m,t+1}$ .

These bounds should limit the predictability in a slightly modified versions of Wright's VAR regressions.<sup>23</sup> Wright's VAR uses a combination of continuously compounded and semiannual yields and inflation compensation, of course, so the bounds do not directly apply to all equations. The bounds should apply to any VAR equation predicting with continuously compounded gross yields (i.e., Treasury yields) because those equations can be transformed linearly into return equations. That is, log gross yields are transformations of log prices— $m \times \ln(1 + y) = -\ln(p)$ —and returns are differenced log prices. One can similarly convert semiannual corporate yields to continuously compounded gross yields. This paper does not transform the inflation compensation variables because they are not asset prices and can take negative values. These transformations produce a new VAR that is very similar to the original VAR.

We can denote the transformed vector as  $\tilde{y}_t$ , where  $\tilde{y}_{i,t} = \ln(1 + y_{i,t})$  for  $i = 1, 2, 5$ , and  $6$  and  $\tilde{y}_{i,t} = y_{i,t}$  for  $i = 3$  and  $4$ .<sup>24</sup> One could write the VAR in transformed data as follows:

$$\tilde{y}_t = \tilde{A} \tilde{y}_{t-1} + \tilde{c} + \varepsilon_t, \quad (6)$$

where  $\tilde{A}$  and  $\tilde{c}$  will be very similar to  $A$  and  $c$  to the extent that the data transformation is linear. Subtracting  $\tilde{y}_{t-1}$  from both sides of (6) then relates the differenced variables—a transformation of returns—to the lagged level variables. The result resembles an error correction framework:

$$\tilde{y}_t - \tilde{y}_{t-1} = -r_t = (\tilde{A} - I)\tilde{y}_{t-1} + \tilde{c} + \varepsilon_t. \quad (7)$$

Equation (7) contains essentially the same information as the original VAR in yields, (1), but is written with some dependent variables as log gross yield differences, i.e., approximate returns.

---

<sup>23</sup> The reader might wonder if rational bounds should apply in the post-crisis financial environment of 2008:11-2011:10. Standard measures of market stress indicate that the market was definitely more volatile than average but still functioning within normal bounds. For example, the mean value of the MOVE index over the volatile sample (2008:11 to 2009:09:30) was higher than the MOVE index in “normal” times (1998-2007) 76 percent of the time.

<sup>24</sup> The transformed data are extremely similar to the original data and difficult to tell apart graphically.

One can estimate (7) to determine if the in-sample  $R^2$ s are too large to be consistent with rational pricing models for a given level of risk aversion. Excessive predictability indicates that the VAR overfits the data. One can also gauge excessive fit by comparing the in-sample  $R^2$  of the regressions in (7) with a Campbell and Thompson (2008) OOS  $R^2$  statistic,

$$R_{OOS}^2 = 1 - \frac{\sum_{t=1}^T (r_t - \hat{r}_t)^2}{\sum_{t=1}^T (r_t - \bar{r}_t)^2}, \quad (8)$$

where  $\hat{r}_t$  is the fitted value from a predictive, OOS regression of returns using expanding sample coefficients and data through  $t-1$  and  $\bar{r}_t$  is the historical average return estimated with data through  $t-1$ . The  $R_{OOS}^2$  statistic—reported in the same units as the in-sample  $R^2$ —measures the proportional reduction in RMSFE for the predictive regression relative to the historical average. A positive value thus indicates that the predictive regression forecasts better than the historical average in terms of RMSFE, while a negative value signals the opposite.

To determine the extent to which the VAR might overfit the data, Table 3 reports the in-sample  $R^2$ , the OOS  $R^2$ , and the Potì and Wang (2010) and Potì and Siddique (2013) bounds on the  $R^2$ s implied the bond return regressions (equation (7)) from the bias-adjusted VAR. The OOS forecast statistics are constructed with expanding samples, updating the VAR coefficients every 20 business days. The  $R^2$  bounds are calculated for relative risk aversions of 2.5 and 5, with a generous estimate of the annualized standard deviation of the market return: 20 percent.<sup>25</sup> Levich and Potì (2015) cite Ross (2005) to argue that 5 is an upper bound on risk aversion.

Table 3 shows clear results: Every in-sample  $R^2$  estimate for bond returns is well above—5 to 7 times as big as—the higher bound on  $R^2$  in a rational pricing model (compare columns 2 and

---

<sup>25</sup> More recent works by Huang and Zhou (2017) and Potì (2018) have tightened the predictability bound by restricting the volatility of the SDF and the return-SDF correlation. It is likely that these methods would further reduce the plausible in-sample  $R^2$  from those presented in Table 3.

5). The predictability would still be too high even if relative risk aversion was as high as 10. The minimum in-sample  $R^2$ s for a bond return is 2.0 percent, for the 10-year Treasury, which is 5 times the 0.4 percent upper bound for daily  $R^2$ . Tellingly, the OOS  $R^2$ s are negative for most of the yield/return regressions but positive for the inflation compensation equations. These negative values are consistent with the VAR's poor OOS forecasting performance. In summary, the VAR has too much in-sample return predictability to be consistent with rational pricing and the negative OOS  $R^2$ s strongly indicate that the in-sample predictability is spurious.

## 6. A VAR That Is Consistent with Rational Pricing

One can restrict the coefficients in the return equations in (7) to produce  $R^2$ s that are consistent with rational asset pricing and then convert the estimated system back to a VAR in yields to obtain impulse response functions and other statistics.<sup>26</sup> One might hope that such a restricted VAR would also be more stable over time than the unrestricted VAR. Of course, this restricted model does not constitute independent evidence for persistence; rather, it formalizes restrictions on persistence implied by rational asset pricing.

We estimate such a bias-corrected VAR on the transformed yield data over the in-sample period in an unrestricted form—equation (6)—and under the restrictions that the  $R^2$ s implied for the bond return equations in (7) cannot exceed the upper bounds in Table 3: 0.004 and 0.001. We then examine the forecasting performance and implied impulse responses of these models. The restricted models were not estimated with a bias correction.<sup>27</sup>

---

<sup>26</sup> One can use Kuhn-Tucker conditions to restrict the VAR coefficients. A binding constraint restricts the return regression coefficients to be proportional to but smaller than the unrestricted coefficients:  $\hat{B} = \left( \frac{1}{k} Y'X(X'X)^{-1}X'Y((Y - \bar{Y})(Y - \bar{Y})')^{-1} \right)^{-1/2} (X'X)^{-1}X'Y$ , where  $Y$  denotes the vector of the dependent variable (i.e., the return),  $X$  denotes the matrix of independent variables, and  $k$  is the upper bound on the  $R^2$ . The restriction effectively shrinks the VAR representation of the coefficients toward an identity matrix.

<sup>27</sup> As the restriction tightens (i.e., as the allowable  $R^2$  goes to zero), the VAR representation would become arbitrarily close to a martingale.

The three panels of Table 4 show the OOS RMSE Theil statistics under an unrestricted VAR and similar VARs restricting the  $R^2$ s to 0.004 and 0.001, respectively.<sup>28</sup> All three VARs used the same transformed data. Although none of the VARs consistently outperform the martingale model in the OOS period for all variables (i.e., the Theil statistics usually exceed 1), restricting the  $R^2$  improves the OOS forecasting performance for bond returns. The most tightly restricted model has the best Theil statistics (bottom panel), and the unrestricted model (top panel) has the worst OOS Theil statistics for those bond returns.

The upper panel of Figure 5 shows that restricting the  $R^2$ s in the bond yield equations to only 0.004 increases the half-life of the shock to the 10-year Treasury bond from 108 days to more than a year. The estimated half-lives of Baa and Aaa yield responses also become much longer, extending from 3-to-5 months to a year or more. Figure 5 omits confidence intervals for clarity, but these tend to become tighter—naturally—as more restrictions are imposed.

This exercise is an application of theory-consistent forecasting. Giacomini and Ragusa (2014) have developed a method to construct multivariate density forecasts that are consistent with moment restrictions implied by economic theory. These methods could have been used to construct the impulse responses in Figure 5. In a manner similar to the way that this paper bounds the kernel volatility to avoiding overfitting bond returns, Kozak, Nagel, and Santosh (2020) bound the volatility of an SDF to price the cross-section of stocks with factors derived from firm characteristics, where theory informs the shrinkage methods. This reduces cross-sectional overfitting. The theoretical motivations preclude near-arbitrage opportunities and assume that major sources of return co-movement drive expected return premia.

---

<sup>28</sup> Note first that the Theil statistics for the unrestricted VAR in log gross yields (top panel of Table 4) are very similar to the Theil statistics for the baseline VAR in net yields in panel 3 of Table 1.

## 7. Alternative Sources of Evidence on Persistence

Would other methods, such as local linear projections (LP), be more reliable than a VAR?

Jorda (2005, 2009) argues for the merits of LP, while Kilian and Kim (2011) conclude that the bias-adjusted bootstrap VAR confidence interval is more accurate than that of LP. Plagborg-Møller and Wolf (2021) establish the equivalence of impulse responses from local linear projections and VARs. While a comparison of empirical methods is beyond the scope of this paper, any empirical method can potentially exhibit structural instability or sensitivity to outliers that render its conclusions suspect. Known risk or liquidity factors, however, could explain findings of strong predictability in asset prices.

Although the VAR’s evidence is unreliable due to the unstable nature of the system of equations, at least three alternative sources of evidence imply persistent UMP effects: Quantity-of-debt models, expectations of professional forecasters, and portfolio/lending behavior.

First, researchers have used “quantity-of-debt” outstanding to predict yields or term premia: Gagnon et al. (2011), Krishnamurthy and Vissing-Jorgensen (2011), Hamilton and Wu (2012), Chadha, Turner, and Zampolli (2013), Greenwood and Vayanos (2014) and Altavilla, Carboni, and Motto (2015). Most such models use the Vayanos-Vila (2021) segmented-markets model, either to motivate regressions or construct formal term-structure models. These studies typically find that asset purchases reduce yields in a manner similar to the lower end of the estimates from event studies.<sup>29</sup> Although quantity-of-debt models do not explicitly estimate the dynamic impact of purchases, they would imply persistent effects of asset purchases because balance sheet effects are themselves very persistent.

---

<sup>29</sup> A weakness of most quantity-of-debt studies is the failure to account for the forward-looking nature of financial markets. For example, a central bank purchase announcement will immediately affect yields/premia, although the bond purchases may not commence for months and holdings might not peak for years. Quantity-of-debt studies also do not account for signaling effects from central bank purchases.

Second, respondents to the Survey of Professional Forecasters expected persistent UMP effects on yields (Altavilla and Giannone, 2017). If UMP shocks were transient, the fact eluded these market professionals.

Third, micro evidence shows that UMP shocks repeatedly and persistently changed portfolio choices, lending, and capital flows. For example, Carpenter et al. (2015) and Joyce, Liu, and Tonks (2017) respectively characterize sectoral and institutional rebalancing in the wake of bond purchases, while Rodnyansky and Darmouni (2017), Kurtzman, Luck, and Zimmermann (2018), and Chakraborty, Goldstein, and MacKinlay (2020) study how Fed asset purchases affected bank lending as a function of bank asset holdings. Intriguingly, McInish, Neely, and Planchon (2021) show that shorts—known to be sophisticated investors—show no evidence of expecting mean reversion after UMP shocks, but rather adjusted their portfolios to benefit from further yield declines (increases) after expansionary (contractionary) shocks.

In summary, quantity-of-debt models, surveys of professional forecasters, and micro studies of lending/portfolio rebalancing imply persistent effects of UMP.

## 8. Discussion and Conclusion

Event studies show that the Federal Reserve’s UMP announcements elicited the desired effects on asset prices and substantially reduced U.S. and foreign long-term yields, as well as the value of the dollar. These immediate, large reductions in long yields were often followed by weeks or months of increases in yields, however, which many observers interpreted to mean that the unconventional shocks had very transient effects on asset prices. That interpretation implies that UMP has limited ability to stimulate the economy.

Wright (2012) offers a clever and potentially very helpful method to measure shock persistence: He identifies a SVAR on interest rate and inflation compensation data under the



assumption that interest rate variance is higher on monetary announcement days. The model implies that UMP shocks have very transient effects on long yields, with median half-lives of perhaps 3 to 6 months. Greenlaw et al. (2018) cite transient effects—among other arguments—in contending that asset purchases had much smaller effect than commonly believed.

The present paper has shown that several serious problems preclude Wright's (2012) SVAR from providing reliable evidence that UMP shocks are transient. The model's conclusions are sensitive to the exclusion of one observation. Omitting March 18, 2009 from the list of announcements essentially removes all evidence of transient shocks. Accounting for lag length uncertainty with model averaging implies substantially more persistence in the 10-year Treasury rate, with the point estimate of the half-life of the shock more than doubling. One should not place too much confidence in this model-average, however, as the SVAR(1) exhibits serious problems. It forecasts very poorly OOS and is structurally unstable, indicating that the dynamic relations that the VAR coefficients purport to describe cannot be trusted and therefore the data do not support the transience of monetary policy shocks. The SVARs with longer lag lengths likely share these problems. These problems highlight the importance of examining structural stability and sensitivity to individual observations in empirical work.

This paper does not argue for any of the alternative estimated models, which are not necessarily consistent with each other. This paper also does not argue that we should eschew SVARs or select models for their forecasting performance. The usefulness of SVARs in answering economic questions outweighs any forecasting advantage of other models.

In addition to considering stability and sensitivity of econometric estimates, this paper argues that researchers should look skeptically at models that imply asset-price predictability that substantially exceeds that consistent with theory. The Wright (2012) SVAR violates bounds on

predictability implied by rational asset pricing models. SVARs that are constrained to be consistent with such models forecast better and generate much more persistent impulse responses to monetary policy shocks. This formalizes the notion that rational asset pricing must imply fairly persistent effects of shocks on asset prices.

The weight of the evidence on UMP shock persistence from other types of studies—quantity-of-debt analyses, micro studies of portfolio choices, and surveys of professional forecasters—suggests that UMP shocks are persistent. However, we cannot precisely measure how persistent the shock effects are.

This study has implications for the monetary policy of non-U.S. central banks. Because the evidence supports persistent effects of Federal Reserve actions, the similar unconventional monetary actions of the ECB, Bank of Japan, Bank of England, Swiss National Bank, Bank of Sweden, and Bank of Denmark probably also had persistent effects. Such persistence is consistent with the findings of Altavilla and Giannone (2017) and Altavilla et al. (2019) and the literature on significant international policy spillovers and portfolio flows, e.g., Neely (2015) and Bauer and Neely (2014).

What might cause the apparent mean reverting behavior of yields after UMP shocks?<sup>30</sup> Wright (2012) suggested two possibilities: 1) that markets simply initially overreacted to the quantitative easing actions and 2) that the stimulus provided by the monetary policy actions caused a delayed increase in yields by stimulating the economy.

With respect to the first explanation, Hamilton (2018) cites an example of overreaction: the stock market's 24-hour plunge and recovery immediately after President Trump's 2016 election. This seems to be an unlikely explanation in the present case. While the 2016 election was a one-

---

<sup>30</sup> Recall that the sensitivity analysis showed that the mean reversion in yields was largely driven by the rise in yields from April through June 2009, following the strong UMP expansion announced on March 18, 2009.

time event, systematically transient UMP reactions would require that many markets—international bond, forex, stock—consistently, irrationally overreacted to UMP announcements over a period of 5-10 years, creating predictable, unexploited profit opportunities. Bhattarai and Neely (forthcoming) discuss this issue.

The second hypothesis is that successful UMP actions sowed the seeds of their own reversal by generating confidence and expectations of growth. According to Wright (2012, p. F464), “A possible—although optimistic—interpretation is that the economic stimulus provided by these Federal Reserve actions caused the economy to pick up.” This “delayed stimulus effect” hypothesis, however, might require too much predictability in long yields to be consistent with rational pricing and reasonable risk aversion.

In addition to the two explanations proposed by Wright, at least two more are possible. A third explanation is that nonmonetary shocks coincidentally increased long yields after the UMP actions. For example, Meyer and Bomfim (2009) argue that higher expected growth, new Treasury issuance, and the return of investors’ risk appetite drove the increase in Treasury yields from late March through mid-June 2009. Rising equity and oil prices over that March-to-June period is consistent with higher expected growth and increasing risk appetites boosting yields.

Finally, a fourth explanation relies on the tendency of the U.S. Treasury to endogenously extended the maturity of U.S. debt in response to UMP yield declines.<sup>31</sup> Greenwood et al. (2014) argue the Treasury’s debt restructuring offset about 1/3 of the cumulative impact of the Fed’s QE policies, as estimated by other studies. Bhattarai and Neely (forthcoming) discuss this.

The latter three interpretations would be consistent with the effectiveness of UMP.

---

<sup>31</sup> The fourth explanation—endogenous Treasury issuance—differs from the Treasury issuance mentioned under reason three in that it is specifically a reaction to yield declines caused by UMP expansions, not simply a coincidental occurrence.

## Appendix A: Bias Adjustment and Bootstrapping

This appendix briefly describes the bias adjustment used in this paper, as well as many previous papers. It follows the discussions in Kilian (1998) and Efron and Tibshirani (1993).

1. Estimate the VAR parameter matrix,  $A$ , with the original  $T \times k$  sample to obtain the OLS estimates of the parameters,  $A_{OLS}$ , residuals,  $\varepsilon_{OLS}$ , and the associated covariance matrix.
2. Using the estimated VAR as the data-generating process, bootstrap 10,000 samples of size  $T \times k$ , by resampling from the residuals,  $\varepsilon_{OLS}$ , using coefficients  $A_{OLS}$  and drawing initial conditions from the unconditional distribution of the data. The residuals were separated into two sets for resampling. The two sets contained residuals from days with and without monetary announcements. Simulated data for non-announcement days were generated with a moving block bootstrap of length 10 from the second set of residuals, and simulated data for announcements were generated by sampling residuals from the first set. This procedure maintained the assumed heteroskedasticity of the data-generating process. Results were robust to variations of the block length or use of the wild bootstrap.
3. Estimate the VAR parameter matrix,  $A^*$ , for each simulated dataset using OLS, and calculate the average of those matrices,  $A_{MC}$ , over the bootstrapped samples.
4. The difference between the true parameters for the simulated data-generating process,  $A_{OLS}$ , and that of the average estimated VAR coefficient matrix,  $A_{MC}$ , is the estimate of the bias in the original VAR on the real data. Therefore, the bias-adjusted coefficient matrix is computed as  $A_{BA} = A_{OLS} + (A_{OLS} - A_{MC})$ .
5. The modulus of  $A_{BA}$  is checked to ensure that it implies a stationary system. If it does not, the bias correction term,  $(A_{OLS} - A_{MC})$ , is gradually reduced until the modulus of the implied  $A_{BA}$  is less than 1.

## Appendix B: The Bound on $R^2$ s in Asset Return Equations

Risk-neutral investors would bid away any positive expected excess returns. Therefore, predictability must stem from risk aversion, that is, the unwillingness of individuals to substitute consumption across states of the world. This appendix summarizes previous work that has established the limits of predictability.

Although Kirby (1998) was important in connecting asset return predictability to risk aversion, this appendix essentially condenses work from Potì and Siddique (2013), Cochrane (1999) and Potì and Wang (2010). Levich and Potì (2015) apply these arguments.

Strictly speaking, the arguments here apply to predicting excess returns, but they can be modified to apply directly to returns. In Wright's (2012) sample—November 2008 through September 2011—the riskless rate was nearly zero and very stable, so excess returns to holding a bond are nearly identical to returns to holding the same bond.

Intuitively, we might think that the degree of predictability should be related to the risk aversion of the marginal investor. Risk-neutral marginal investors, for example, should bid away any predictable excess returns, but the predictable component of returns could grow with the risk aversion of the marginal investor, as he is less willing to bid prices to their risk-neutral level.

To formalize this notion, first recall the usual stochastic discount factor (SDF) relation for excess returns, where  $m_{t+1}$  is the SDF and  $r_{t+1}^e$  is an excess return.

$$E(m_{t+1}r_{t+1}^e|I_t) = 0 \quad (\text{B.1})$$

The definition of covariance between the SDF and excess returns is as follows:

$$\text{Cov}(m_{t+1}r_{t+1}^e|I_t) = E(m_{t+1}r_{t+1}^e|I_t) - E(m_{t+1}|I_t) E(r_{t+1}^e|I_t). \quad (\text{B.2})$$

Using this definition, (B.2) and the properties of the SDF in (B.1), we get

$$\text{Cov}(m_{t+1}r_{t+1}^e|I_t) = -E(m_{t+1}|I_t) E(r_{t+1}^e|I_t)$$

Appendices are not for publication.

$$\Rightarrow E(r_{t+1}^e | I_t) = -\frac{\text{Cov}(m_{t+1}r_{t+1}^e | I_t)}{E(m_{t+1} | I_t)} = -(1 + i_t)\text{Cov}(m_{t+1}r_{t+1}^e | I_t) \quad (\text{B.3})$$

where  $(1 + i_t)$  is the gross, one-period riskless rate at time  $t$  and the second equality uses the well-known fact that  $E(m_{t+1} | I_t) = \frac{1}{(1+i_t)}$ . Because correlations must be less than or equal to one in absolute value, i.e., because the Cauchy–Schwarz inequality must hold,

$$\text{Cov}(r_{t+1}^e, m_{t+1} | I_t) \leq \sigma(r_{t+1}^e)\sigma(m_{t+1}) . \quad (\text{B.4})$$

Using (B.3) and (B.4), we have

$$-E(r_{t+1}^e | I_t) = (1 + i_t)\text{Cov}(r_{t+1}^e, m_{t+1} | I_t) \leq (1 + i_t) \sigma(r_{t+1}^e)\sigma(m_{t+1}) \quad (\text{B.5})$$

Squaring both sides, we get

$$E(r_{t+1}^e | I_t)^2 \equiv E(\mu_t^2) \leq (1 + i_t)^2 \sigma^2(r_{t+1}^e)\sigma^2(m_{t+1}) \quad (\text{B.6})$$

Dividing by  $\sigma^2(r_{t+1}^e)$  and using the definition of  $R^2$  as the quotient of the variance in expected variation over total variance of excess returns, implies a bound on predictability, i.e., the  $R^2$ , in terms of the variance of the SDF.

$$R^2 \equiv \frac{E(\mu_t^2)}{\sigma^2(r_{t+1}^e)} \leq (1 + i_t)^2 \sigma^2(m_{t+1}) \quad (\text{B.7})$$

Potì and Wang (2010) approximate the SDF around current wealth with a Taylor series that explicitly uses the intertemporal marginal rate of substitution (IMRS) as the SDF. See equations (5)-(7) in Potì and Wang (2010). The first order approximation expresses the SDF in terms of the first and second derivatives of a time-separable utility function of initial market wealth and the market return,  $r_{m,t+1}$ .

$$m_{t+1} \cong \frac{U''(W_{m,t})}{U'(W_{m,t})} W_{m,t} r_{m,t+1} = \text{RRA } r_{m,t+1} \quad (\text{B.8})$$

That expresses the variance of the SDF in terms of the generic utility function, wealth and market return. The second equality on the right-hand-side of (B.8) follows from the definition of the coefficient of relative risk aversion (RRA) in terms of derivatives of the utility function.

Appendices are not for publication.

$$\sigma_t^2(m_{t+1}) \cong \sigma^2(RRA\ r_{m,t+1}) = RRA^2\ \sigma^2(r_{m,t+1}) \quad (\text{B.9})$$

Using (B.9) in (B.7) while conservatively setting  $1 + i_t = 1$ —because gross interest rates are typically larger than 1—implies that the product of the squared coefficient of relative risk aversion and the variance of the market return bounds the  $R^2$  for a predictive regression of excess returns.

$$R^2 \leq (1 + i_t)^2 RRA^2\ \sigma^2(r_{m,t+1}) \leq RRA^2\ \sigma^2(r_{m,t+1}) \quad (\text{B.10})$$

The remainder of this appendix briefly describes the related approach of Kirby (1998), closely following that paper's exposition. Kirby (1998) starts with (B.1) but normalizes the SDF to have expectation of one by dividing it by its own mean:  $\tilde{m}_t = \frac{m_t}{E(m_t)}$ . Such a normalized SDF is clearly still valid for excess returns.

$$E(\tilde{m}_{t+1}r_{t+1}^e|I_t) = 0 \quad (\text{B.11})$$

For any element of the information set,  $z_t \in I_t$ , it must be the case that

$$E(\tilde{m}_{t+1}r_{t+1}^ez_t|I_t) = 0. \quad (\text{B.12})$$

The law of iterated expectations implies that this is also true unconditionally,

$$E(\tilde{m}_{t+1}r_{t+1}^ez_t) = 0 \quad . \quad (\text{B.13})$$

Using (B.11), the definition of covariance, and the fact that  $E(\tilde{m}_t)=1$  by construction,

$$E(r_{t+1}^e) = -cov(\tilde{m}_{t+1}, r_{t+1}^e). \quad (\text{B.14})$$

(B.14) is essentially the first equality in (B.3). (B.14) also holds for the excess return to any dynamic strategy return:  $z_t * r_{t+1}^e$ .

$$E(r_{t+1}^e * z_t) = -cov(\tilde{m}_{t+1}, r_{t+1}^e * z_t). \quad (\text{B.15})$$

The first equality below uses the definition of covariance, while the second uses (B.15), the third uses (B.14) and the fourth uses the properties of expectations.

Appendices are not for publication.

$$\begin{aligned}
cov(r_{t+1}^e, z_t) &= E(r_{t+1}^e * z_t) - E(r_{t+1}^e) * E(z_t) \\
&= -cov(\tilde{m}_{t+1}, r_{t+1}^e * z_t) - E(r_{t+1}^e) * E(z_t) \\
&= -cov(\tilde{m}_{t+1}, r_{t+1}^e * z_t) + cov(\tilde{m}_{t+1}, r_{t+1}^e) * E(z_t) \\
&= -cov(\tilde{m}_{t+1}, r_{t+1}^e * (z_t - E(z_t)))
\end{aligned} \tag{B.16}$$

Finally, the definition of covariance implies

$$\begin{aligned}
&cov(\tilde{m}_{t+1}, r_{t+1}^e * (z_t - E(z_t))) = \\
&E(\tilde{m}_{t+1} * r_{t+1}^e * (z_t - E(z_t))) - E(\tilde{m}_{t+1}) * E(r_{t+1}^e * (z_t - E(z_t)))
\end{aligned} \tag{B.17}$$

Again using (B.13) and  $E(\tilde{m}_t)=1$ , we get

$$cov(\tilde{m}_{t+1}, r_{t+1}^e * (z_t - E(z_t))) = -E(r_{t+1}^e * (z_t - E(z_t))) \tag{B.18}$$

So, using (B.18) in (B.16), we get

$$cov(r_{t+1}^e, z_t) = E(r_{t+1}^e * (z_t - E(z_t))) \tag{B.19}$$

Kirby (1998) considers the restrictions that such a model would impose in the context of a linear regression of an excess return,  $r_{t+1}^e$ , on a constant and a  $k \times 1$  vector of non-deterministic regressors,  $z_t$ , where the  $k+1$  coefficients are partitioned into a vector,  $b = [b_0, b_z']'$ , corresponding to the partition of the regressors,  $x_t = [1, z_t']'$ .

The population  $R^2$  in this regression is the explained sum of squares over the total sum of squares or

$$R^2 = \frac{b_z' \Sigma_{zz} b_z}{\sigma_r^2}, \tag{B.20}$$

where  $\Sigma_{zz}$  is the covariance matrix of the non-deterministic regressors,  $z_t$ , and  $\sigma_r^2$  is the variance of the excess returns. The first equality in (B.21) is from the usual OLS formula while the second equality relies on a trivial modification of (B.15).

$$b = E[x_t x_t']^{-1} E[r_{t+1}^e * x_t] = -E[x_t x_t']^{-1} cov(\tilde{m}_{t+1}, r_{t+1}^e * x_t) \tag{B.21}$$



Appendices are not for publication.

Using (B.18), (B.19) and extracting the coefficients on the variable regressors from (B.21), we get an expression for the coefficients,  $b_z$ , on the variable regressors in terms of the SDF's covariance with the product of the excess return and the instrument,  $z_t$ .

$$b_z = \Sigma_{zz}^{-1} \text{cov} \left( \tilde{m}_{t+1}, r_{t+1}^e * (z_t - E(z_t)) \right) \quad (\text{B.22})$$

Kirby (1998) goes on to use this framework to test whether the predictive ability for monthly stock portfolios is consistent with rational pricing under several candidate SDFs, including several consumption models and several CAPM specifications. That is, Kirby asks if the regression coefficients implied by the models differ significantly from those generated by unrestricted least squares. Although the returns implied by size factors are too predictable, Kirby (1998) deemed the overall predictability to be consistent with rational asset pricing.

## References

- Altavilla, C., Brugnolini, L., Gürkaynak, R., Motto, R., and Ragusa, G. (2019). “Measuring Euro Area Monetary Policy,” *Journal of Monetary Economics*, vol. 108, pp. 162–79.
- Altavilla, C., Carboni, G., and Motto, R. (2015). “Asset Purchase Programmes and Financial Markets: Lessons from the Euro Area.” European Central Bank Working Paper 1864.
- Altavilla, C. and Giannone, D. (2017). “The Effectiveness of Non-Standard Monetary Policy Measures: Evidence from Survey Data,” *Journal of Applied Econometrics*, vol. 32, pp. 952–964.
- Andrews, D. (1993). “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, vol. 61, pp. 821–56.
- Bhattarai, S. and Neely, C. (forthcoming). “An Analysis of the Literature on International Unconventional Monetary Policy,” *Journal of Economic Literature*.
- Bai, J. and Perron, P. (1998). “Estimating and Testing Linear Models with Multiple Structural Changes,” *Econometrica*, vol. 66, pp. 47–78.
- Bai, J. and Perron, P. (2003). “Computation and Analysis of Multiple Structural Change Models,” *Journal of Applied Econometrics*, vol. 18, pp. 1–22.
- Bauer, M. and Neely, C. (2014). “International Channels of the Fed’s Unconventional Monetary Policy,” *Journal of International Money and Finance*, vol. 44, pp. 24–46.
- Baumeister, C. and Benati, L. (2013). “Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound,” *International Journal of Central Banking*, vol. 9, pp. 165–212.
- Bekaert, G. and Hodrick, R. (1992). “Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets,” *Journal of Finance*, vol. 47, pp. 467–509.
- Belke, A., Gros, D. and Osowski, T. (2017). “The Effectiveness of the Fed’s Quantitative Easing Policy: New Evidence Based on International Interest Rate Differentials,” *Journal of International Money and Finance*, vol. 73, pp. 335–349.
- Bernanke, B. (2012). “Opening Remarks: Monetary Policy Since the Onset of the Crisis,” In *Proceedings - Economic Policy Symposium - Jackson Hole*, pp. 1–22.
- Bernanke, B. (2020). “The New Tools of Monetary Policy,” *American Economic Review*, vol. 110, pp. 943–83.
- Buckland, S., Burnham, K. and Augustin, N. (1997). “Model Selection: An Integral Part of Inference,” *Biometrics*, vol. 53, pp. 603–618.

- Campbell, J. and Thompson, S. (2008). “Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?” *Review of Financial Studies*, vol. 21, pp. 1509–531.
- Carpenter, S., Demiralp, S., Ihrig, J., and Klee, E. (2015). “Analyzing Federal Reserve Asset Purchases: From Whom Does the Fed Buy?” *Journal of Banking and Finance*, vol. 52, pp. 230-44.
- Chadha, J., Turner, P., and Zampolli, F. (2013). “The Interest Rate Effects of Government Debt Maturity.” BIS Working Paper 415.
- Chakraborty, I., Goldstein, I., and MacKinlay, A. (2020). “Monetary Stimulus and Bank Lending,” *Journal of Financial Economics*, vol. 136, pp. 189-218.
- Cochrane, J. (1999). “Portfolio Advice for a Multifactor World,” Federal Reserve Bank of Chicago *Economic Perspectives*, vol. 23, pp. 59–78.
- Cochrane, J. (2007). “The Dog That Did Not Bark: A Defense of Return Predictability,” *The Review of Financial Studies*, vol. 21, pp. 1533-1575.
- Cochrane, J. and Piazzesi, M. (2005). “Bond Risk Premia,” *American Economic Review*, vol. 95, pp. 138-160.
- Duffee, G. (2013). “Forecasting Interest Rates,” in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2A, Elsevier Publications, pp. 385–426.
- Duffie, D. (2010). “Presidential Address: Asset Price Dynamics with Slow-Moving Capital,” *Journal of Finance*, vol. 65, pp. 1237–67.
- Efron, B. and Tibshirani, R. (1993). “An Introduction to the Bootstrap,” *Monographs on Statistics and Applied Probability*, vol. 57, New York and London: Chapman and Hall.
- Faust, J., Rogers, J. and Wright, J. (2003). “Exchange Rate Forecasting: The Errors We’ve Really Made,” *Journal of International Economics*, vol. 60, pp. 35–59.
- Feunou, B., and Fontaine, J. (2018). “Bond Risk Premia and Gaussian Term Structure Models,” *Management Science*, vol. 64, pp. 1413-39.
- Gabriel, S.A. and Lutz, C. (2017). “The Impact of Unconventional Monetary Policy on Real Estate Markets,” Available at SSRN 2493873.
- Gagnon, J., Raskin, M., Remache, J. and Sack, B. (2011). “The Financial Market Effects of the Federal Reserve’s Large-Scale Asset Purchases,” *International Journal of Central Banking*, vol. 7, pp. 3–43.
- Gagnon, J. (2016). “Quantitative Easing: An Underappreciated Success,” Peterson Institute for International Economics Policy Brief No. PB16-4.

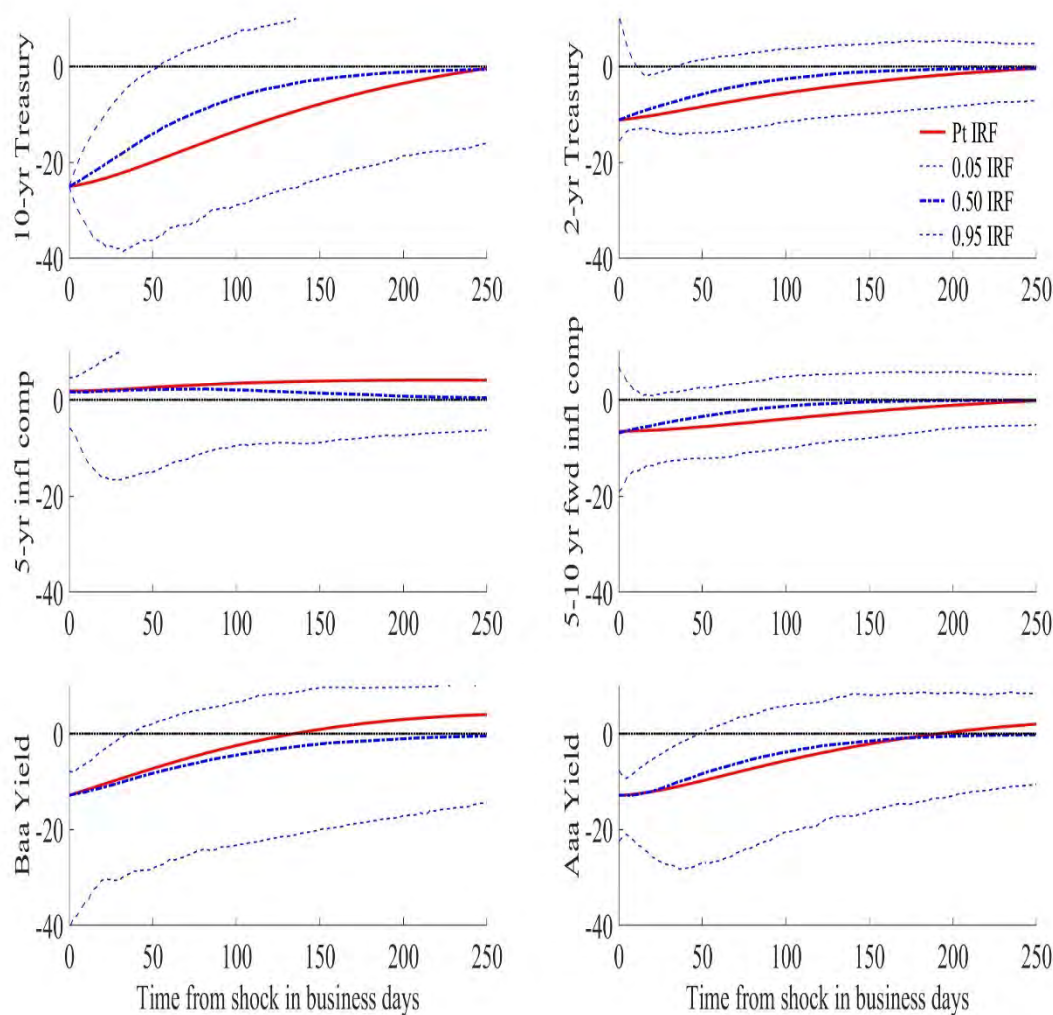
- Gambacorta, L., Hofmann, B. and Peersman, G. (2014). “The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis,” *Journal of Money, Credit, and Banking*, vol. 46, pp. 615-42.
- Gertler, M. and Karadi, P. (2013). “QE 1 vs. 2 vs. 3...: A Framework for Analyzing Large-Scale Asset Purchases as a Monetary Policy Tool,” *International Journal of Central Banking*, vol. 9, pp. 5–53.
- Giacomini, R., and Ragusa, G. (2014). “Theory-Coherent Forecasting,” *Journal of Econometrics*, vol. 182, pp. 145-155.
- Gilchrist, S. and Zakrajšek, E. (2013). “The Impact of the Federal Reserve’s Large-Scale Asset Purchase Programs on Corporate Credit Risk,” *Journal of Money, Credit, and Banking*, vol. 45, pp. 29-57.
- Greenlaw, D., Hamilton, J., Harris, E., and West, K. (2018). “A Skeptical View of the Impact of the Fed’s Balance Sheet,” National Bureau of Economic Research Working Paper 24687.
- Greenwood, R., Hanson, S., Rudolph, J. and Summers, L. (2014). “Government Debt Management at the Zero Lower Bound,” Brookings Institution Hutchins Center on Fiscal and Monetary Policy Working Paper 5.
- Greenwood, R. and Vayanos, D., (2014). “Bond Supply and Excess Bond Returns,” *The Review of Financial Studies*, vol. 27, pp. 663-713.
- Hamilton, J. (2018). “The Efficacy of Large-Scale Asset Purchases When the Short-Term Interest Rate Is at Its Effective Lower Bound,” *Brookings Papers on Economic Activity*, pp. 543–54.
- Hamilton, J. and Wu, C. (2012). “The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment,” *Journal of Money, Credit, and Banking*, vol. 44, pp. 3-46.
- Hansen, S. and McMahon, M. (2016). “Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication,” *Journal of International Economics*, vol. 99, pp. S114-S133.
- Huang, D., and Zhou, G. (2017). “Upper bounds on return predictability,” *Journal of Financial and Quantitative Analysis*, vol. 52, pp. 401-425.
- Jordà, Ò. (2005). “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, vol. 95, pp. 161-182.
- Jordà, Ò. (2009). “Simultaneous Confidence Regions for Impulse Responses,” *The Review of Economics and Statistics*, vol. 91, pp. 629-647.

- Joslin, S., Le, A. and Singleton, K. (2013). “Why Gaussian Macro-finance Term Structure Models are (Nearly) Unconstrained Factor-VARs,” *Journal of Financial Economics*, vol. 109, pp. 604-22.
- Joyce, M., Lasaoa, A., Stevens, I. and Tong, M. (2011). “The Financial Market Impact of Quantitative Easing in the United Kingdom,” *International Journal of Central Banking*, vol. 7, pp. 113–61.
- Joyce, M., Liu, Z., and Tonks, I. (2017). “Institutional Investors and the QE Portfolio Balance Channel,” *Journal of Money, Credit, and Banking*, vol. 49, pp. 1225-46.
- Joyce, M., Miles, D., Scott, A. and Vayanos, D. (2012). “Quantitative Easing and Unconventional Monetary Policy – An Introduction,” *The Economic Journal*, vol. 122, pp. F271-F288.
- Kilian, L. (1998). “Small-Sample Confidence Intervals for Impulse Response Functions,” *Review of Economics and Statistics*, vol. 80, pp. 218–30.
- Kilian, L. and Kim, Y.J., (2011). “How Reliable are Local Projection Estimators of Impulse Responses?” *Review of Economics and Statistics*, vol. 93, pp.1460-1466.
- Kirby, C. (1998). “The Restrictions on Predictability Implied by Rational Asset Pricing Models,” *Review of Financial Studies*, vol. 11, pp. 343–82.
- Kozak, S., Nagel, S, and Santosh, S. (2020). “Shrinking the Cross-Section,” *Journal of Financial Economics*, vol. 135, pp. 271-292.
- Krishnamurthy, A., and Vissing-Jorgensen, A. (2011). “The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy.” *Brookings Papers on Economic Activity*, pp.215-65.
- Kurtzman, R., Luck, S, and Zimmermann, T. (2018). “Did QE Lead Banks to Relax Their Lending Standards? Evidence from the Federal Reserve’s LSAPs.” *Journal of Banking and Finance*, 105403.
- Kuttner, K. (2018). “Outside the Box: Unconventional Monetary Policy in the Great Recession and Beyond,” *Journal of Economic Perspectives*, vol. 32, pp. 121–46.
- Levich, R. and Potì, V. (2015). “Predictability and ‘Good Deals’ in Currency Markets,” *International Journal of Forecasting*, vol. 31, pp. 454–72.
- Litterman, R. (1986). “Forecasting with Bayesian Vector Autoregressions—Five Years of Experience,” *Journal of Business and Economic Statistics*, vol. 4, pp. 25–38.
- Mamaysky, H. (2018). “The Time Horizon of Price Responses to Quantitative Easing,” *Journal of Banking & Finance*, vol. 90, pp. 32-49.

- Manski, C. (1993). "Identification Problems in the Social Sciences," *Sociological Methodology*, vol 23, pp. 1-56.
- Manski, C. (2010). "Partial Identification in Econometrics," in Steven N. Durlauf, and Lawrence E. Blume, (eds.), *Microeconometrics*, Palgrave Macmillan.
- Mark, N. (1995). "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review*, vol. 85, pp. 201–18.
- McInish, T., Neely, C. and Planchon, J. (2021). "Unconventional Monetary Policy and the Behavior of Shorts," Federal Reserve Bank of St. Louis Working Paper 2017-31.
- Meese, R. and Rogoff, K. (1983). "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, vol. 14, pp. 3–24.
- Meyer, L. and Bomfim, A. (2009). "Were Treasury Purchases Effective? Don't Just Focus on Treasury Yields...", Monetary Policy Insights: Fixed Income Focus, Macroeconomic Advisers Report.
- Neely, C. (2015). "Unconventional Monetary Policy Had Large International Effects," *Journal of Banking & Finance*, vol. 52, pp. 101–11.
- Neely, C. and Weller, P. (2000). "Predictability in International Asset Returns: A Reexamination," *The Journal of Financial and Quantitative Analysis*, vol. 35, pp. 601–20.
- Newey, W. and West. K. (1994). "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, vol. 61, pp. 631–53.
- Plagborg-Møller, M. and Wolf, C., (2021). "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, vol. 89, pp. 955-980.
- Poti, V.. (2018). "A New Tight and General Bound on Return Predictability," *Economics Letters*, vol. 162, pp. 140-145.
- Poti, V. and Siddique, A. (2013). "What Drives Currency Predictability?" *Journal of International Money and Finance*, vol. 36, pp. 86-106.
- Poti, V. and Wang, D. (2010). "The Coskewness Puzzle," *Journal of Banking & Finance*, vol. 34, pp. 1827–38.
- Rapach, D. and Wohar, M. (2006). "In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining," *Journal of Empirical Finance*, vol. 13, pp. 231–47.
- Rigobon, R. and Sack, B. (2004). "The Impact of Monetary Policy on Asset Prices," *Journal of Monetary Economics*, vol. 51, pp. 1553–75.

- Rodnyansky, A. and Darmouni, O. (2017). “The Effects of Quantitative Easing on Bank Lending Behavior,” *Review of Financial Studies*, vol. 30 , pp. 3858-87.
- Ross, S. (2005). *Neoclassical Finance*, Princeton, NJ: Princeton University Press.
- Rossi, B. (2013). “Advances in Forecasting Under Model Instability,” in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2B, Elsevier Publications, pp. 1203-324.
- Stock, J. and Watson, M. (2003). “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature*, vol. 41, pp. 788–829.
- Swanson, E. (2021). “Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets,” *Journal of Monetary Economics*, vol. 118, pp. 32-53.
- Thornton, D. and Valente, G. (2012). “Out-of-Sample Predictions of Bond Excess Returns and Forward Rates: An Asset Allocation Perspective,” *Review of Financial Studies*, vol. 25, pp. 3141–68.
- Vayanos, D., and Vila, J. (2021). “A Preferred-Habitat Model of the Term Structure of Interest Rates.” *Econometrica*, vol. 89, pp. 77-112.
- Wang, H., Zhang, X. and Zou, G. (2009). “Frequentist Model Averaging Estimation: A Review,” *Journal of Systems Science and Complexity*, 22, 732-748.
- Welch, I. and Goyal, A. (2008). “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction,” *Review of Financial Studies*, vol. 21, pp. 1455–508.
- Wright, J. (2012). “What Does Monetary Policy Do to Long-Term Interest Rates at the Zero Lower Bound?” *The Economic Journal*, vol. 122, pp. F447–66.
- Yu, E. (2016). “Did Quantitative Easing Work?” Federal Reserve Bank of Philadelphia *Economic Insights*, vol. 1, pp. 5-13.

Figure 1: Impulse responses from the baseline VAR

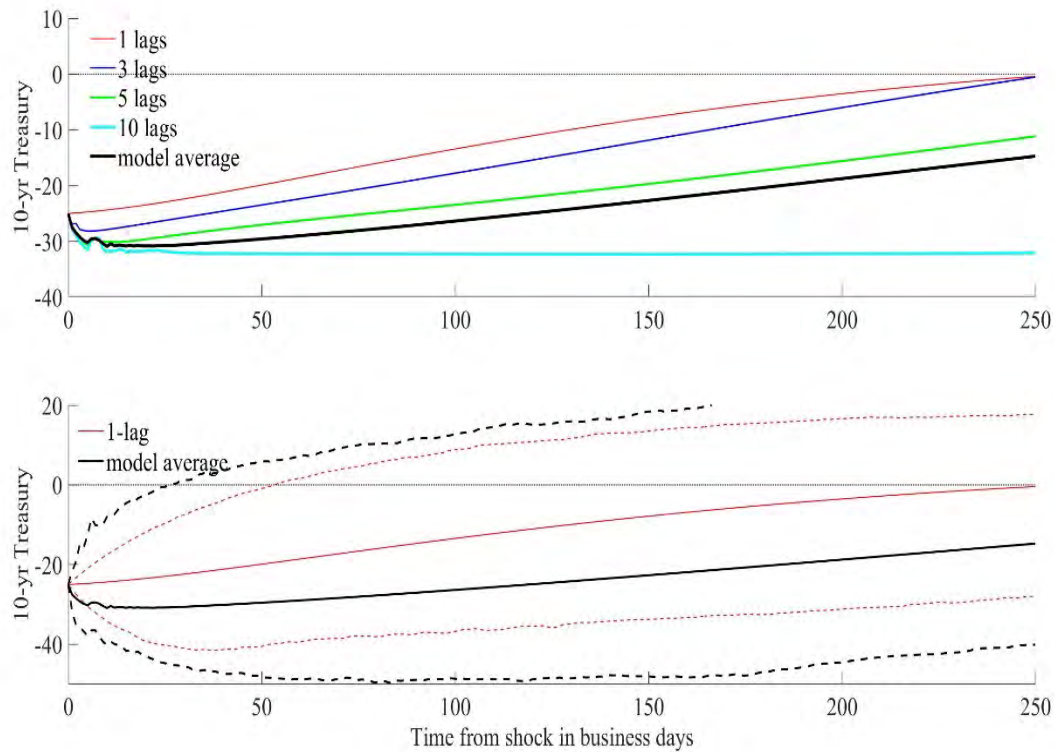


Notes: The figure illustrates impulse responses and 90 percent confidence intervals for the impact of monetary policy shocks on daily yields/inflation compensation in a 6-variable VAR in net yields. The horizontal axes are in days, while the vertical axes are in basis points. The impulse responses are structurally identified by the greater variance of interest rates on days of monetary policy announcements. The figure essentially replicates Wright's (2012) study of the impact of UMP shocks.

Source: Haver Analytics.

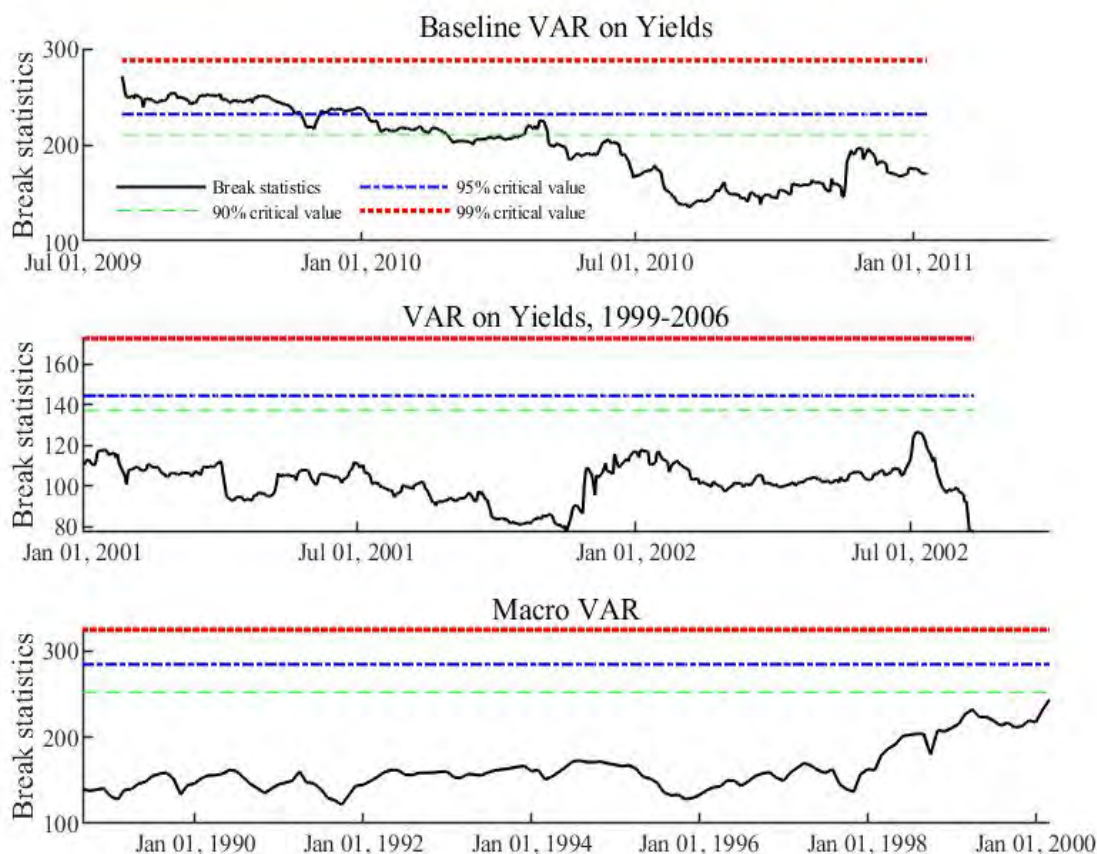


Figure 2: Impulse responses from the baseline VAR with alternative lag lengths



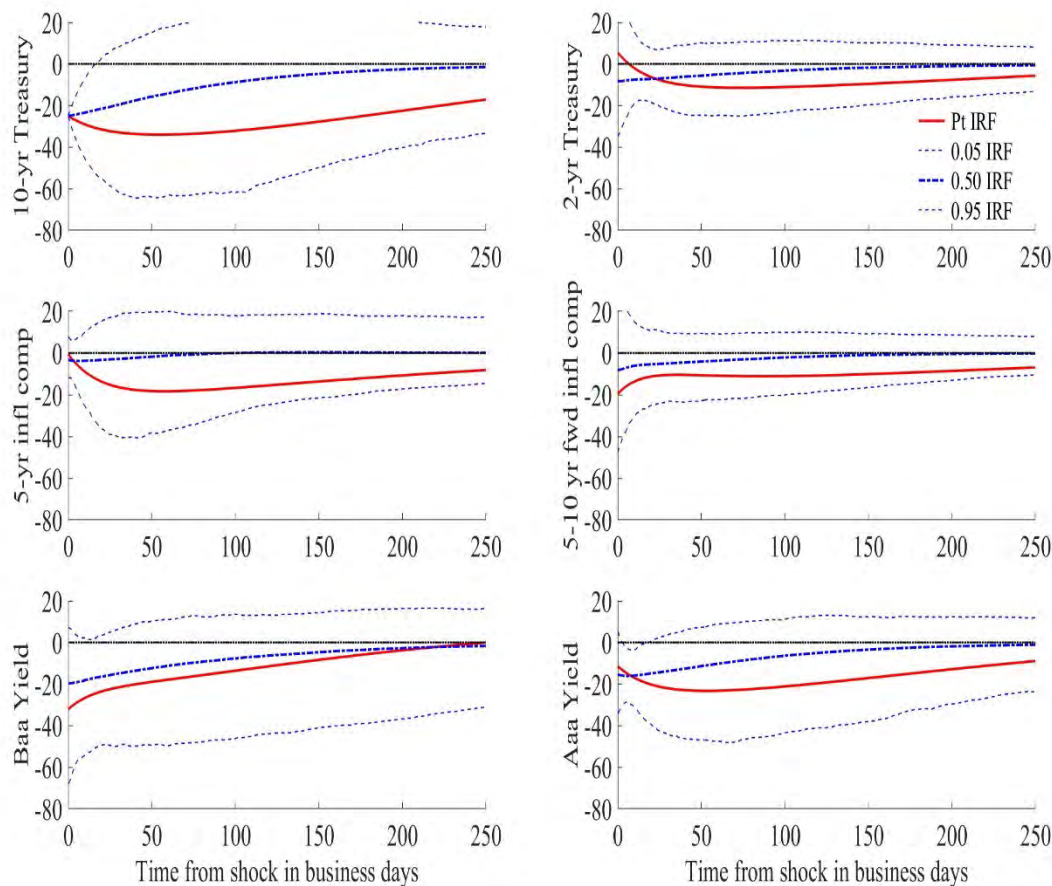
Notes: The top panel of the figure illustrates the responses in basis points of yields on the 10-year Treasury to UMP shocks using VARs with 1, 3, 5 and 10 lags, as well as a frequentist model-average of VARs with lags from 1 to 15. The bottom panel of the figure shows the 1-lag (red) and model averaged (black) impulse responses of the 10-year Treasury along with a bootstrapped 90 percent confidence interval. The confidence intervals on the model-averaged specification are “jagged” because the model is drawing from a mixture of distributions.

Figure 3: Andrews (1993) structural break statistics for the VAR



Notes: The three panels of the figure plot the Andrews (1993) structural break test statistics and the bootstrapped 1, 5, and 10 percent critical values from the 25<sup>th</sup> to the 75<sup>th</sup> percentile of the samples, for three VARs: 1) the baseline VAR in yields, estimated from November 2008 through September 2011; 2) the same VAR in yields, but estimated from 1999 through 2006; and 3) a VAR estimated on macro variables (industrial production, CPI-U, personal consumption expenditures, 3-month Treasury yield, price of West Texas Intermediate crude, and the civilian unemployment rate) from 1983 through 2006. Critical values were obtained with a moving block bootstrap with a 10-day block, overlaid with draws from the days of monetary policy announcements from the distribution of residuals on those days.

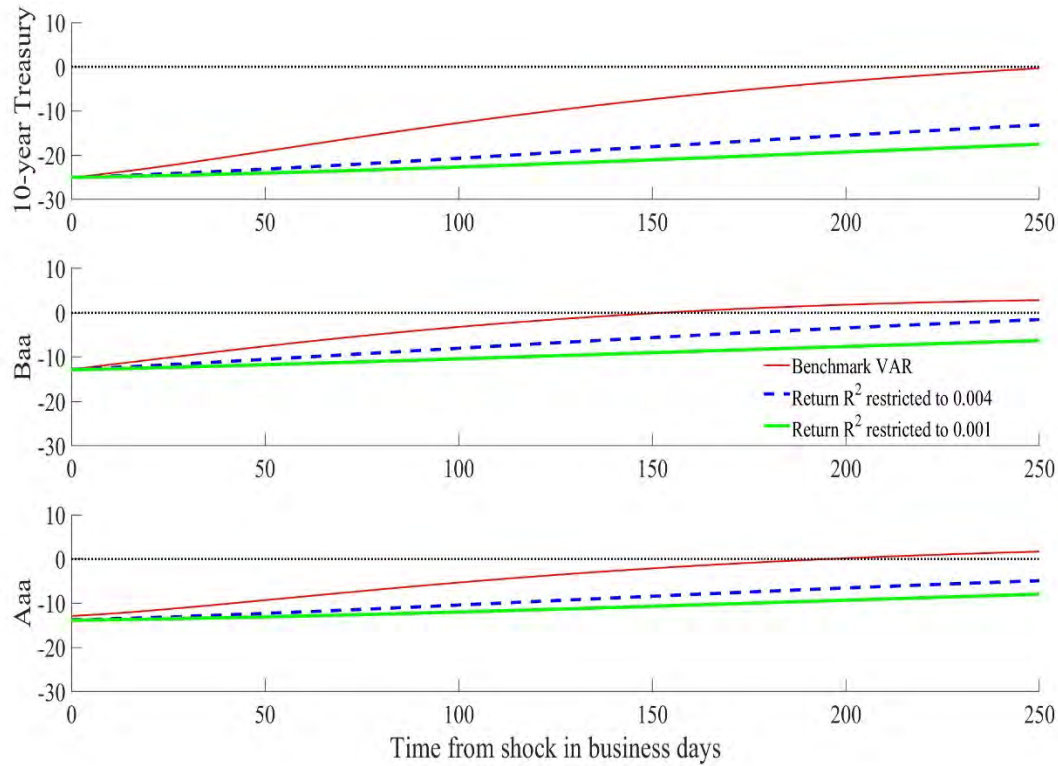
Figure 4: Impulse responses from the VAR with March 18, 2009 reclassified



Notes: The figure illustrates impulse responses and 90 percent confidence intervals for the impact of monetary policy shocks on daily yields/inflation compensation in a 6-variable VAR in net yields, with the date March 18, 2009 omitted from the list of announcement dates. See the notes to Figure 1 for additional information.

Source: Haver Analytics.

Figure 5: Impulse responses from the unrestricted and  $R^2$ -restricted VARs



Notes: The figure illustrates impulse responses in basis points to yields on the 10-year Treasury, Baa, and Aaa bonds from monetary policy shocks. Each panel of the figure illustrates impulse responses from the unrestricted, baseline VAR, and VARs in which the equations for the bond yields—Treasury and corporates—are restricted to imply  $R^2$ s for the returns that do not exceed 0.004 and 0.001, respectively. The structural shocks are identified, as in Wright (2012), by the greater variance of interest rates on days of monetary policy announcements.

Table 1: Out-of-sample root mean squared error forecast statistics

	2-yr Treasury	10-yr Treasury	5-yr infl comp	5-10-yr fwd infl comp	Baa Yield	Aaa Yield
1-day Naïve RMSFE	0.02	0.05	0.03	0.04	0.05	0.05
20-day Naïve RMSFE	0.06	0.22	0.15	0.13	0.19	0.18
60-day Naïve RMSFE	0.07	0.40	0.30	0.18	0.31	0.32
120-day Naïve RMSFE	0.08	0.54	0.34	0.22	0.43	0.48
1-day VAR RMSFE	0.02	0.06	0.03	0.04	0.05	0.06
20-day VAR RMSFE	0.16	0.49	0.15	0.14	0.27	0.41
60-day VAR RMSFE	0.23	0.92	0.28	0.23	0.53	0.75
120-day VAR RMSFE	0.24	1.11	0.27	0.27	0.70	0.91
1-day Theil	1.24	1.12	1.00	0.99	1.01	1.10
20-day Theil	2.63	2.24	0.97	1.02	1.42	2.32
60-day Theil	3.20	2.28	0.93	1.25	1.71	2.39
120-day Theil	2.90	2.04	0.80	1.22	1.62	1.90
1-day Theil $p$ -value	0.01	0.19	0.91	0.96	0.79	0.53
20-day Theil $p$ -value	0.02	0.12	0.88	0.74	0.71	0.46
60-day Theil $p$ -value	0.05	0.21	0.84	0.55	0.70	0.48
120-day Theil $p$ -value	0.17	0.33	0.86	0.56	0.72	0.56

Notes: This table shows OOS root mean square forecast statistics from the 6-variable VAR. The top panel shows the RMSFE for the naïve (martingale) forecast; the second panel shows the RMSFE for the bias-adjusted VAR predictions; the third panel shows Theil statistics, the ratio of the VAR RMSFE to the naïve RMSFE; the fourth panel shows the proportion of samples in which bootstrapped Theil statistics from the VAR null are greater than the actual Theil statistics in the third panel. The OOS period is from October 2011 through November 2013.

Table 2: Bai-Perron structural stability tests

Test	2-yr US Treasury yield	10-yr US Treasury yield	5-yr US infl comp	5-10 yr US fwd infl comp	Moody's Baa yield	Moody's Aaa yield
<i>UD max</i>	<b>29.26</b>	<b>32.21</b>	<b>31.63</b>	<b>21.52</b>	<b>23.38</b>	<b>22.46</b>
<i>WD max</i>	<b>34.40</b>	<b>32.21</b>	<b>32.45</b>	23.25	<b>32.04</b>	30.80
SupF(3 2)	10.74	12.08	10.32	13.23	15.61	15.17
SupF(2 1)	<b>23.62</b>	16.00	<b>22.00</b>	9.30	<b>26.48</b>	<b>27.79</b>

Notes: The table displays the results of Bai-Perron (1998, 2003) tests for breaks at unknown points over the in-sample period. The null hypotheses for the *UD max* and *WD max* tests are that each equation is stable. The null hypothesis for the SupF(3|2) test is that there are no more than 2 breaks in the respective equations; the null hypothesis for the SupF(2|1) test is that there is no more than 1 break in the respective equations. Italicized, bold, and bold/italicized test statistics denote significance at the 10, 5, and 1 percent levels, respectively.

Table 3: In-sample and out-of-sample  $R^2$  as well as rational bounds

	In-sample $R^2$	OOS $R^2$	Rational $R^2$ bounds	
	Bias-corrected	Bias-corrected	RRA = 2.5	RRA = 5
2-yr Treasury return	2.2	-13.1	0.1	0.4
10-yr Treasury return	2.0	-6.0	0.1	0.4
5-yr $\Delta$ inflation comp	1.2	0.6	0.1	0.4
5-yr-fwd $\Delta$ inflation comp	2.1	1.1	0.1	0.4
Baa return	2.3	-3.3	0.1	0.4
Aaa return	2.7	-6.9	0.1	0.4

Notes: This table shows the in-sample  $R^2$ s and Campbell and Thompson (2008) OOS  $R^2$ s, in percentage terms, for the regressions of log returns onto lagged gross log yields (equation (7)) and the bounds on return  $R^2$ s implied by rational asset pricing and given coefficients of relative risk aversion (RRA). These  $R^2$ s are generated with bias-corrected coefficients, as in the baseline VAR. The bounds assume a 20 percent annualized standard deviation for the market return. The coefficients for the expanding, OOS forecasts were updated every 20 days. The in-sample period was November 2008 through September 2011 and the OOS period was October 2011 through November 2013.

Table 4: Out-of-sample root mean square forecast statistics for the restricted VAR

	2-yr Treasury	10-yr Treasury	5-yr infl comp	5-10-yr fwd infl comp	Baa yield	Aaa yield
Bias-corrected VAR						
1-day Theil	1.24	1.12	1.00	0.99	1.01	1.10
20-day Theil	2.62	2.22	<b>0.97</b>	1.02	1.40	2.30
60-day Theil	3.17	2.26	<b>0.93</b>	1.24	1.68	2.36
120-day Theil	2.85	2.02	<b>0.80</b>	1.20	1.60	1.88
$R^2$ restricted to 0.004						
1-day Theil	1.06	1.02	1.00	0.99	1.00	1.01
20-day Theil	1.72	1.34	1.03	0.97	1.05	1.36
60-day Theil	2.51	1.49	1.00	1.03	1.12	1.53
120-day Theil	2.25	1.40	0.84	0.91	<b>1.05</b>	1.33
$R^2$ restricted to 0.001						
1-day Theil	<b>1.01</b>	<b>1.00</b>	<b>1.00</b>	<b>0.99</b>	<b>1.00</b>	<b>1.00</b>
20-day Theil	<b>1.18</b>	<b>1.08</b>	1.09	<b>0.96</b>	<b>1.02</b>	<b>1.06</b>
60-day Theil	<b>1.74</b>	<b>1.16</b>	1.18	<b>0.97</b>	<b>1.09</b>	<b>1.14</b>
120-day Theil	<b>1.98</b>	<b>1.16</b>	0.99	<b>0.85</b>	1.13	<b>1.10</b>

Notes: This table shows OOS root mean square forecast statistics from the 6-variable transformed (equation (6)) VAR. The top panel shows the Theil statistics for the bias-corrected VAR forecast; the second and third panels show the Theil statistics for the VAR predictions with the  $R^2$  restricted to be no greater than 0.004 and 0.001, respectively. Bold Theil statistics show the best of the three models for each horizon-variable combination. The in-sample period was November 2008 through September 2011 and the OOS period was October 2011 through November 2013.