Research Division
Federal Reserve Bank of St. Louis
*Working Paper Series*

# How Persistent Are Unconventional Monetary Policy Effects?

**Christopher J. Neely**

Working Paper 2014-004C
http://research.stlouisfed.org/wp/2014/2014-004.pdf

October 2016

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

# How Persistent Are Unconventional Monetary Policy Effects?

Christopher J. Neely*

This version: October 28, 2016

**Abstract**

Event studies show that the Federal Reserve's announcements of forward guidance and large-scale asset purchases had large and desired effects on asset prices but these studies do not tell us how long such effects last. Wright (2012) used a structural vector autoregression (SVAR) to argue that unconventional policies have very transient effects on bond yields, with half-lives of 3 to 6 months. The present paper shows, however, that this inference is unsupported for several reasons. First, accounting for model uncertainty greatly lengthens the estimated persistence. Second, and more seriously, the inference is unreliable because the SVAR is structurally unstable and forecasts very poorly. Finally, the implied in-sample return predictability from the SVAR greatly exceeds a level consistent with rational asset pricing and reasonable risk aversion. Restricted models that respect more plausible asset return predictability are more stable and imply that unconventional monetary policy shocks were fairly persistent. 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; phone: +1-314-444-8568; fax: +1-314-444-8731. Christopher J. Neely is an assistant 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, 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 and Evans Karson 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 created special facilities to support lending, expanded swap lines with foreign central banks, 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 pursued outright asset purchases to support the economy, especially housing markets. These quantitative easing (QE) purchases occurred in several 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. Collectively, these programs committed the Fed to purchasing trillions of dollars of long-term assets.

Event studies have established that QE announcements strongly affected asset prices. Gagnon et al.'s (2011a,b) study finds 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. Hamilton and Wu (2012) indirectly calculate the effects of the Fed's 2008-09 QE programs with a term structure model. Neely (2015) evaluates the effect of QE1 on international long bond yields and exchange rates. Bauer and Neely (2014) evaluate the relative importance of the signaling and portfolio balance channels on international interest rates with term structure models.

Despite this profusion of event studies on purchase announcements, there has been much less work on the impact of QE on macroeconomic variables (Baumeister and Benati, 2010; Gambacorta, Hofmann, and Peersman, 2012; 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 very difficult, however, to estimate the persistence of unconventional policy 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. The heteroskedasticity in interest rates on days of unconventional monetary policy shocks identifies the contemporaneous effects of unconventional monetary shocks (Rigobon and Sack, 2004, and Craine and Martin, 2008). Wright's impulse responses imply that unconventional monetary policy shocks have large, but very transient, effects on U.S. interest rates; most of the impact of large-scale asset purchases (LSAPs) on 10-year Treasury, Aaa-rated, and Baa-rated yields dissipate within 6 months. This is consistent with anecdotal conclusions. Many market observers concluded that QE1 failed because long yields rose in the late spring of 2009 (Woodhill, 2013). Figure 1 illustrates this more than 100-basis-point rise from late April through June.

There are good reasons to question this finding, however. Wright's methods assume that a VAR accurately describes the stable dynamics of the first two moments of the data, but overfitting and structural instability are ubiquitous in models of asset prices. Meese and Rogoff (1983) made this point forcefully in the context of structural models of the exchange rate. They showed that very poor out-of-sample (OOS) performance accompanied good in-sample 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 vintage. Goyal and Welch (2008) question the usefulness of traditional equity premium predictors in OOS forecasting exercises. Ubiquitously poor OOS forecasting reflects the structural instability of the equations explaining asset returns. The estimated dynamic relations are spurious.

The contribution of this paper is its careful analysis of the VAR system used by Wright to show that the data do not support the inference that unconventional monetary policy shocks have transient effects.[1] Specifically, the VAR lag length is likely misspecified, the VAR forecasts very poorly OOS, and fails structural stability tests. Therefore, its conclusions about shock persistence are unreliable. In contrast, a naive, no-change model outperforms the VAR, suggesting that monetary shocks likely have very persistent effects.

This paper also argues that transient policy effects are inconsistent with rational asset pricing and reasonable risk aversion because the former would create an opportunity for risk-adjusted expected returns. Restricted VAR models that are consistent with reasonable risk aversion and rational asset pricing forecast better than the unrestricted VAR and imply more plausible, persistent responses to monetary shocks.

This paper confronts the specific problem of the persistence of unconventional monetary policy shocks in the context of a VAR, but it also has a much larger lesson: Exercises to estimate the dynamic path of asset prices are to be viewed with caution. Thus, the evidence supports the view that unconventional monetary policy shocks probably have fairly persistent effects on long yields but we cannot tell exactly how persistent, and our uncertainty about the effects of shocks grows with the forecast horizon.

The next section of the paper describes Wright's SVAR methodology. Section 3 describes and replicates Wright's main findings. Section 4 illustrates the failure of VAR forecasting and structural stability. Section 5 shows that the baseline VAR is inconsistent with rational asset

[1] This paper does not examine all of Wright's conclusions. In a second part of his paper, Wright also constructs a set of unconventional monetary shocks from high-frequency data. This paper does not critique that methodology.

pricing and reasonable risk aversion, while Section 6 presents the results of restricted VARs that are consistent with rational pricing and reasonable risk aversion. Section 7 concludes.

## 2. The Structural VAR Methodology

The reduced-form VAR can be written as follows:

$$A(L)y_t = \varepsilon_t, \tag{1}$$

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

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

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 $\left(I - A(L)\right)^{-1} R_i$.[2]

Ordinary least squares (OLS) estimates of coefficients on lagged endogenous regressors will be biased in finite samples and will generally underestimate the persistence of the data. Therefore, Wright follows Kilian (1998) in correcting this bias with a bootstrapping procedure.[3]

---

[2] 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.

[3] Mankiw and Shapiro (1986) and Stambaugh (1986) discuss the small-sample bias imparted by lagged endogenous regressors. Bekaert, Hodrick, and Marshall (1997) use Monte Carlo procedures to correct for such a bias in term structure tests. Although the present paper constructs bias-corrected impulse responses and forecasts, uncorrected models behave grossly similarly.

Bauer, Rudebusch, and Wu (2014) criticize the lack of bias correction in term structure models in Wright (2011). Wright (2014) responds that the uncorrected estimates are more consistent with survey measures.

For the present paper, the author experimented with several types of bootstrapping, all of which incorporated the assumed heteroskedastic structure of the data-generating process. The results presented here use a bootstrap drawn from two distributions, a moving block bootstrap, with the block length of 10, for non-monetary policy days,

Appendix A describes the bias correction and bootstrapping methods used in this paper.

The identification scheme assumes neither a pattern of contemporaneous interactions (Bernanke, 1986; Blanchard and Watson, 1986; and Sims, 1986) nor long-run relations (Shapiro and Watson, 1988, and Blanchard and Quah, 1989). Instead, 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.[4]

Wright assumes only that the variance of the structural monetary policy shock, $u_{t,1}$, is higher on 28 specific monetary announcement days ($o_{1,A}^2$) than on non-announcement days ($o_{1,N}^2$), which creates heteroskedastic reduced-form errors, $\varepsilon_t$.[5] The announcement set included dates of Federal Open Market Committee (FOMC) meetings and other announcements or speeches by the Chairman that were relevant to unconventional monetary policy.[6] 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:

$$\Sigma_1 - \Sigma_0 = R_1 R_1'\left(o_{1,A}^2 - o_{1,N}^2\right). \tag{3}$$

The VAR estimation of (1) provides estimates of $A(L)$, $\Sigma_1$, and $\Sigma_0$ that enable one to estimate $R_1$ from (3). Because the terms in the product $R_1 R_1'\left(o_{1,A}^2 - o_{1,N}^2\right)$ are not separately identified, Wright normalizes $\left(o_{1,A}^2 - o_{1,N}^2\right)$ to 1 and solves for the elements of $R_1$ by minimizing the

overlaid with draws on the monetary policy days from the distribution of residuals on those days. Results were fairly similar with other bootstrapping methods (e.g., the wild bootstrap or other block lengths).

[4] Rudebusch (1998a) marks the first use of financial market data to describe monetary shocks in a VAR. Sims (1998) criticizes this approach and Rudebusch (1998b) responds.

[5] 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.

[6] 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.

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 response functions for the unconventional monetary policy shocks. As Wright is interested only in the impact of unconventional monetary policy shocks, there is no need for additional identifying assumptions.

Wright 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., $R_i R_i' = (\Sigma_1 - \Sigma_0)$). 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-year, 10-year forward inflation compensation yields; and the Moody's Baa- and Aaa-rated corporate bond yield indices—using daily data from November 3, 2008, to September 30, 2011. The Moody's yields are semiannually compounded; the Treasury yields are continuously compounded. The block bootstrap provides confidence intervals on the impulse response functions. This paper replicates Wright's VAR results with similar data, estimation 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:

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

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 2 illustrates the resulting impulse responses and 90 percent bootstrapped confidence intervals, which are similar to — perhaps a bit wider than — those in Wright's paper.[7] Monetary policy shocks significantly change 10-year Treasury, Baa, and Aaa rates, with the corporate rates showing immediate effects ranging from 40 to more than 100 percent of the Treasury changes.[8] This immediate effect is consistent with event studies of unconventional monetary policy (e.g., Gagnon et al., 2011a,b, and Neely, 2015).

The significant initial effects of the monetary policy shock wear off very quickly, however. The half-lives of the responses of the 10-year Treasury and corporate yields range from 3 to 6 months. Although the 90 percent confidence interval for the 10-year Treasury indicates that one cannot reject that the half-life of the monetary shock on that yield is 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, Miles, Scott, and Vayanos (2012), Gagnon (2016), and Yu (2016). The existence of only short-lived effects is a potentially very important result: It suggests that unconventional policy actions have only very transient effects on yields and therefore very modest effects on macroeconomic

[7] The results presented use a 10-day moving block bootstrap overlaid with separate draws for the announcement days. The results are not very sensitive to the block length or the use of the wild bootstrap.
[8] The bias-corrected coefficients that produced Figure 2 do imply a stationary system, but the implied behavior of the response point estimates is 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.

variables. Wright (2012, p. F465) summarizes as follows: "*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 impulse response confidence intervals in Figure 2 imply fairly precise estimates of dynamic behavior. But these confidence intervals are both pointwise—that is, narrower than uniform confidence bands—and conditional on the assumed lag length for the VAR. 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.[9] Full results are omitted for brevity but are available on request.

[9] 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.

Philosophically, this model selection exercise points out a decision theoretic problem with standard econometric practice. Applied econometricians typically search for a parsimonious model, setting parameters equal to zero that are statistically insignificant (i.e., small compared with the precision of the estimate). The pragmatic desire to avoid overfitting, poor OOS forecasting, and spurious economic inference motivates this practice. 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. The assumption that imprecisely 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 the conditioning imparted by model selection tests, it is worth examining impulse responses generated by VARs with longer lag lengths. The top panel of Figure 3 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.[10] 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. We shall see that including these "small" coefficients significantly

Further, it is well-known that standard term structure models fit the cross-section well but dynamics badly. There is evidence that additional lags (Cochrane and Piazzesi (2005), Joslin et al., (2013)) or moving average terms (Feunou and Fontaine (2015)) can improve the dynamic fitting.

[10] Although the VAR(10) impulse responses appear to show potential nonstationary behavior, examination of very long-horizon behavior confirms that the system is stationary.

increases the estimated persistence in the VAR, however.

The fact that the economic inference depends on the number of lags in the VAR 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 alone is very likely to distort inference toward choosing a lower lag length and inferring transient shocks. Full results are available from the author.

This exercise does not reveal that the BIC is a bad model selection tool. Given an infinite amount of data from a stable data generating process, it 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. But a parsimonious model will not necessarily describe the long-run dynamics well. Rather, this exercise suggests that the apparent transience of the monetary policy shocks in the VAR(1) is partly due to the emphasis on point estimates and may be an artifact of the lag length selection process, which is heavily influenced by the short length of the sample.

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 model's BIC, as suggested by Buckland, Burnham, and Augustin (1997).[11] The top panel of Figure 3 illustrates the point

[11] Wang, X Zhang, G Zou (2009) usefully review the literature on frequentist model averaging.

estimate of the impulse response of the 10-year Treasury from the averaged estimated, using a model set of VARs from 1 to 15 lags. The bottom panel of Figure 3 displays the point estimates and 90 percent confidence intervals from the 1-lag and averaged model. The averaged model implies a much more persistent response than does the 1-lag model and its confidence interval is shifted toward persistence. The half-life of the shock is more than doubled and one cannot reject the hypothesis of no diminution in the shock for more than a year.[12] In other words, formally accounting for model uncertainty substantially increases the estimated persistence of the monetary policy shocks on the 10-year Treasury. The next sections of the paper, however, suggest that the VAR(1) exhibits more serious problems that imply great caution about drawing conclusions on persistence from VAR models.

## 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 2 are spurious and the data fail to support their apparent implication that unconventional policy has very transient effects. That is, although unrestricted VARs are not necessarily the best forecasting models, they nevertheless must describe stable dynamic relations between the variables to accurately describe the responses of variables to shocks.

A potentially serious difficulty is that 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 a sundry asset prices in OOS exercises: exchange rates (Meese and Rogoff, 1983; Faust, Rogers, and Wright, 2003); equities (Goyal and Welch, 2008); interest rates (Thornton and Valente, 2012), and cross-asset studies (Neely and Weller, 2000).

[12] 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.

### *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.

The in-sample forecasting results are unremarkable and are omitted for brevity. Within the estimation sample, the VARs have lower root mean square forecasting errors (RMSFEs) than the naive, no-change predictions and are approximately unbiased at horizons from 1 to 120 days.

Even excellent in-sample performance does not necessarily mean that the model will predict OOS returns better than some simple benchmark model. A well-specified VAR with a stable covariance structure should be able to forecast asset prices during the OOS period, 2011-13, with the parameter estimates from 2008-11. Thus, OOS forecasting implicitly tests the structural stability of the VAR structure.

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.[13]

[13] The overlapping $n$-period forecast errors will have at least an $n$–1-order serial correlation that must be taken into account in the tests.

Table 1 shows the OOS RMSFEs in basis points, over 1-, 20-, 60- and 120-day horizons, for a naive, 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 naive 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 naive, no-change prediction is superior to the VAR forecasts for 18 of 24 horizon-yield combinations considered (Table 1). The only cases for which the VAR is competitive with the no-change forecast are for changes in inflation compensation. Even for these variables, the VAR does not clearly outperform the naive forecast.

In contrast, the naive, no-change forecast outperforms the VAR at every horizon for every yield variable. This strongly suggests that the VAR does not accurately model true dynamic relations between the variables and that the martingale describes them better. In other words, the impulse response functions in Figure 2 are very likely based on spurious dynamic relations.

Table 2 shows the OOS bias (mean errors) and Newey-West *t*-statistics for the null of unbiased forecasts (Newey and West, 1994). The naive predictions are never systematically biased in a statistically significant way but the VAR yield predictions are biased at all horizons.

It is true that misspecified models sometimes forecast better than correctly specified models, out-of-sample. This, however, is generally only true when the correctly specified model's parameters are estimated so poorly in a finite sample that the misspecified model actually describes the dynamic relations better than the correctly specified, but poorly estimated, model. A correctly specified linear model with the true parameters will always outforecast a misspecified model. In the present case, the naïve model clearly outperforms the VAR(1) in 5 of

6 equations at nearly every horizon, indicating that the estimated VAR(1) describes the dynamics very poorly.

One might think that modifying the VAR procedure would improve the forecasting performance and rescue the possibility of constructing informative impulse responses for monetary shocks. There is considerable evidence that combining Bayesian techniques with VARs is helpful in forecasting (Litterman, 1986). Wright (2012), however, already considered such techniques in his robustness checks and found impulse responses that are similar to those in Figure 2, which suggests that the VARs that produced them also have unstable moments. Of course, one could tighten up the priors on the Bayesian VAR to essentially reproduce the naive forecasts, but then one would obtain very persistent impulse responses, not the mean-reverting impulse responses that indicate transient effects.

### *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 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.[14] 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.

[14] We construct the Andrews test statistics for the uncorrected VAR coefficients to keep computational cost within reasonable bounds. The bias correction—the addition of a very "small" matrix to the VAR coefficients—is very unlikely to change the outcome of the structural stability tests. In addition, the asymptotic test statistics for structural stability should be identical for both sets of VAR coefficients.

Consistent with this poor OOS forecasting performance, the top panel of Figure 4 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 2 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. 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.[15]

This paper follows those Bai-Perron guidelines under the assumptions of a maximum of 3 breaks with at least 20 percent of the original sample between each break. The first and second rows of Table 3 show that the *UD max* and *WD max* statistics reject the null of no breaks for all equations at conventional significance levels. The third row of Table 3 illustrates that one 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. In summary, all equations exhibit at least one break and most exhibit at least two breaks.

Figure 4 and Table 3 are strong evidence against stability: 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.

[15] 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).

Such tests typically require large separation between the models, i.e., big breaks.

In summary, the VAR that produced Figure 2—evidence for the transient effects of unconventional monetary policy shocks—forecasts very poorly OOS and fails tests of structural stability, for both the whole VAR and the individual equations. That is, the data do not support the inference from Figure 2 that monetary shocks are transient. Instead, the relative success of the martingale model indicates that very persistent shocks probably better approximate the dynamic structure.

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

The highly significant break statistics in the top panel of Figure 4 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 more calm 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 moving block bootstrap to simulate data. The center and bottom panels of Figure 4 display the break statistics for the two 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 2008:11–2011:10 sample were likely an important factor in the instability.

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

In a world of risk neutral investors, expected excess returns should be bid to zero. But non-zero expected excess returns are consistent with risk-averse investors and greater risk aversion should permit more 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.

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.[16]

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

where $R_f$ is the riskless rate, $RRA_V$ is the upper bound on relative risk aversion (RRA), and $\sigma^2\left(r_{m,t+1}\right)$ is the variance of the market excess return, $r_{m,t+1}$.

Are the VAR relations that Wright estimates consistent with these rational bounds?[17] Wright's VAR uses a combination of continuously compounded and semiannual yields and inflation compensation, of course, so the bounds don't 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, The inflation compensation spreads are not

[16] 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.
[17] 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 sample (2008:11 to 2009:09:30) was higher than the daily MOVE index in "normal" times (1998- 2007) 76 percent of the time.

exactly liquid asset prices and can take negative values. Therefore, this paper does not transform the inflation compensation variables. 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 $\tilde{y}_{i,t} = \ln\left(1 + y_{i,t}\right)$ for *i* = 1, 2, 5, and 6 and $\tilde{y}_{i,t} = y_{i,t}$ for *i* = 3 and 4.[18] One could write the VAR in transformed data as follows:

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

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:

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

The advantage of (7) is that it contains essentially the same information as the original VAR in yields, (1), but it enables us to judge the plausibility of the in-sample fit versus a rational asset pricing model for those four equations within (7) —the four yield equations—that can be written with the dependent variable as the difference of log gross yields, i.e., returns.

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,

$$R_{OS}^2 = 1 - \frac{\sum_{t=1}^{T}(r_t - \widehat{r_t})^2}{\sum_{t=1}^{T}(r_t - \overline{r_t})^2}, \tag{8}$$

where $\widehat{r_t}$ is the fitted value from a predictive, OOS regression of returns using expanding sample

[18] The transformed data are extremely similar to the original data and difficult to tell apart graphically.

coefficients and data through *t–1* and $\overline{r_t}$ is the historical average return estimated with data through *t–1*. The $R^2_{OS}$ 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 outperforms 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 4 reports the in-sample $R^2$, the OOS $R^2$, and the bounds on the $R^2$s implied by Kirby's (1998) calculations on the bond return data (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 values of relative risk aversion of 2.5 and 5, with a generous estimate of the annualized standard deviation of the market return: 20 percent. Levich and Potì (2015) cite Ross (2005) to argue that 5 is an upper bound on reasonable values of risk aversion.

Table 4 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 (columns 2 and 5). 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 and smaller 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.[19] 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 4: 0.001 and 0.004. We then examine the forecasting performance and implied impulse responses of these models. The restricted models were not estimated with a bias correction.[20]

The three panels of Table 5 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.[21] All three VARs used the same transformed data. Although none of the VARs consistently outperform the martingale model in the OOS period (i.e., the Theil statistics usually exceed 1), restricting the $R^2$ improves the OOS forecasting performance. The most tightly restricted model has the best Theil statistics (bottom panel), and the unrestricted model (top panel) has the worst OOS Theil statistics. This pattern is clearest for the bond yield equations.

Figure 5 illustrates the greater shock persistence implied by restricted models. In particular, the upper panel of Figure 5 shows that restricting the $R^2$s in the bond yield equations to 0.004

---

[19] 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: $\hat{B} = \left(\frac{1}{k} Y'X(X'X)^{-1}X'Y\left((Y-\bar{Y})'(Y-\bar{Y})\right)^{-1}\right)^{-1/2} (X'X)^{-1}X'Y$, 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.

[20] As the restriction tightens (i.e., as the allowable $R^2$ goes to zero), the VAR representation would become arbitrarily close to a martingale.

[21] Note first that the Theil statistics for the unrestricted VAR in log gross yields (top panel of Table 5) are very similar to the Theil statistics for the baseline VAR in net yields in panel 3 of Table 1.

increases the half-life of the shock to the 10-year Treasury bond from 109 days to 241 days. Restricting the $R^2$s to 0.001 increases the half-life of the point estimate of the shock to the 10-year bond to more than a year. The point estimates for the Aaa and Baa bonds similarly get much more persistent. The confidence intervals are omitted for clarity in the graph but tend to become tighter—naturally—as more restrictions are imposed. One cannot reject the hypothesis that the half-life of the shock exceeds a year for any case.

## 7. Discussion and Conclusion

Event studies show that the Federal Reserve's unconventional monetary policy 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. Many observers interpreted these rising yields in the wake of QE announcements to mean that the unconventional shocks had very transient effects on asset prices. If that interpretation were true, it would suggest that unconventional policy has very limited ability to stimulate the economy.

It is very difficult, however, to measure the persistence of the monetary shock effects on yields. Wright (2012) offers a clever and potentially very helpful resolution to this problem: He identifies a structural VAR on interest rate and inflation compensation data under the assumption that interest rate variance is higher on monetary announcement days. The impulse response functions from this VAR(1) indicate that unconventional monetary shocks have very transient effects on long yields, with median half-lives of perhaps 3 to 6 months.

The present paper showed that accounting for lag length uncertainty with model averaging implies substantially more persistence in the estimated response of the 10-year Treasury rate, with the half-life of the shock more than doubling. One should not place too much confidence in this model-average, however, as the VAR(1) exhibits serious problems that suggest caution in drawing any conclusions about shock persistence from a VAR estimated on this sample. In particular, the VAR(1) forecasts very poorly, OOS, and is structurally unstable. Structural instability implies that the dynamic relations that the VAR coefficients purport to describe do not really exist and therefore the data do not support the transience of monetary policy shocks shown in Figure 2.

In addition, the estimated VAR violates bounds on predictability in rational asset pricing models. VARs that are constrained to be consistent with rational asset pricing 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. We cannot measure, however, precisely how persistent any such effects are.

Although this paper has confronted a specific problem—the duration of the effects of monetary policy shocks in a VAR—it has a larger lesson: One should be circumspect in forecasting asset prices or describing their dynamics. There is good reason to believe that asset prices should exhibit limited predictability and the empirical literature has repeatedly confirmed this point.

This paper does not argue that we should discard SVARs or select models on their forecasting performance. Structural VARs can usefully answer interesting economic questions that outweigh the fact that other models forecast better. But economists should respect the limited predictability of asset prices in estimating dynamic relations. Further, standard model selection procedures can affect inference and asset pricing models are often unstable.

How should one interpret the rise in yields after expansionary unconventional monetary policy shocks? Wright suggested the first two possibilities: 1) that the stimulus provided by the monetary policy actions caused a delayed increase in yields by stimulating the economy and 2) that markets simply initially overreacted to the quantitative easing actions. 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. Another interpretation is that markets initially overreacted to the news of these quantitative easing actions.*

The first hypothesis is that successful unconventional monetary policy actions sowed the seeds of their own reversal by generating higher confidence and expectations of higher growth. This "delayed stimulus effect" hypothesis, however, would require a lot of predictability in long yields that is probably inconsistent with rational pricing and reasonable risk aversion.

The second hypothesis is that markets simply overreacted. Such an "overreaction" interpretation would require systematic overreaction to many FOMC announcements over a period of almost five years, through the "taper tantrum" of the summer and fall of 2013. It seems implausible that market participants fail to learn from repeated mistakes.

A third explanation is that all sorts of shocks continually influence the economy and asset prices and that nonmonetary shocks coincidentally increased long yields after the unconventional policy 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. A parallel rise in equity and oil prices over the same March-to-June period is consistent with the explanation that higher expected growth and a rise in appetite for risk raised long rates. This "other shocks" interpretation would be consistent with the usual assumption that shocks to asset prices are very persistent.

Appendices are not for publication.

**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 fairly 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.

Appendices are not for publication.

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

In a risk-neutral environment, investors would bid away any positive expected excess returns. Therefore, predictability must stem from risk aversion, 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 (1997) was important in connecting asset return predictability to risk aversion, this appendix essentially condenses work from the appendices to Potì and Siddique (2013), who draw on arguments in Cochrane (1999) and Potì and Wang (2010). Their arguments are applied in Levich and Potì (2015).

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.

Researchers often predict excess returns with a set of conditioning variables, *X*:

$$r_t = \beta X_t + \varepsilon_t \tag{B.1}$$

Potì and Siddique (2013) point out that the $R^2$ of such a regression can be written in terms of the variance of the return's predictable component, $\mu_t$, to its total variance, $\sigma_\mu^2$:

$$R^2 \equiv \frac{\sigma_\mu^2}{\sigma_r^2} = \frac{E(\mu_t^2) - E(\mu_t)^2}{\sigma_r^2} = \frac{E(\mu_t^2)}{\sigma_\mu^2/1 - R^2} - \frac{E(\mu_t)^2}{\sigma_r^2}$$

$$= E\left(\frac{\mu_t^2}{\sigma_\mu^2}\right)(1 - R^2) - E\left(\frac{\mu_t}{\sigma_r}\right)^2 = E\left(\left(\frac{\mu_t}{\sigma_\mu}\right)^2\right)(1 - R^2) - SR(r_t)^2, \tag{B.2}$$

where $E\left[\left(\frac{\mu_t}{\sigma_u}\right)^2\right]$ is the squared <u>conditional</u> Sharpe ratio (SR) and $E\left(\frac{\mu_t}{\sigma_r}\right)$ is the SR to a static long position. Introducing the notation $SR_{t-1}(r_t)$ to denote a conditional SR, (A.2) becomes

Appendices are not for publication.

$$R^2 = E(SR_{t-1}(r_t)^2)(1 - R^2) - SR(r_t)^2. \tag{B.3}$$

Potì and Siddique (2013) cite Cochrane (1999) to argue that the squared unconditional SR is the expectation of the squared conditional SR, $SR(r_t)^2 = E(SR_{t-1}(r_t)^2)$, and they denote the return to a time-varying position in the asset as $r_t^*$. The unconditional SR of such a strategy can be denoted as $SR(r_t^*)^2 = E(SR_{t-1}(r_t^*)^2) = E(SR_{t-1}(r_t)^2)$ and one can rewrite (A.3) as

$$R^2 = SR(r_t^*)^2(1 - R^2) - SR^2(r_t). \tag{B.4}$$

Solving (B.4) for the squared unconditional SR of the time-varying strategy,

$$SR(r_t^*)^2 = \frac{SR^2(r_t) + R^2}{1 - R^2}, \tag{B.5}$$

Potì and Siddique (2013) interpret the SR of the time-varying strategy on the left-hand side as a function of the SR of a "static" long position in the asset—$SR(r_t)$—and the coefficient of determination $R^2$ of the predictive strategy. Equation (B.5) implies that the $R^2$ of the predictive regression cannot exceed the squared SR of the time-varying investment strategy:

$$R^2 \leq SR(r_t^*)^2. \tag{B.6}$$

Citing Potì and Wang (2010) and Ross (2005), Potì and Siddique (2013) use a capital asset pricing model framework to bound the SR of the time-varying trading strategy as a function of the relative risk aversion (RRA) of the marginal trader and the market return:

$$SR(r_t^*)^2 \leq RRA_V^2 \sigma(r_{m,t})^2, \tag{B.7}$$

where $r_{m,t}$ is the return to the market portfolio at time *t*. Combining (B.6) and (B.7), we obtain the bound on the $R^2$ for the predictive regression as a function of the RRA of the marginal trader and the variance of the market return, $\sigma^2(r_{m,t+1})$:

$$R^2 \leq \left(1 + R_f\right)^2 RRA_V^2 \sigma^2(r_{m,t+1}) \cong RRA_V^2 \sigma^2(r_{m,t+1}). \tag{B.8}$$

## References


Andrews, D. (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica,* vol. 61, pp. 821–56.

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.

Bauer, M.D., Rudebusch, G.D. and Wu, J.C. (2014). "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset: Comment," *American Economic Review*, vol. 104(1), pp. 323–37.

Baumeister, C. and Benati, L. (2010). "Unconventional Monetary Policy and the Great Recession," European Central Bank Working Paper no. 1258.

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.

Bekaert, G., Hodrick, R. and Marshall, D. (1997). "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics*, vol. 44, pp. 309–48.

Bernanke, B. (1986). "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, vol. 25(1), pp. 49–99.

Blanchard, O. and Quah, D. (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, vol. 79(4), pp. 655–73.

Blanchard, O. and Watson, M. (1986). "Are Business Cycles All Alike?" in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press.

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: an integral part of inference. Biometrics, 603-618.

Campbell, J. and Thompson, S. (2008). "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, vol. 21, pp. 1509–531.

Cochrane, J. (1999). "Portfolio Advice for a Multifactor World," Federal Reserve Bank of Chicago *Economic Perspectives*, vol. 23(3), pp. 59–78.

Cochrane, J. H., and Piazzesi, M. (2005). "Bond risk premia." *The American Economic Review*, vol. 95(1), pp. 138-160.

Craine, R. and Martin, V. (2008). "International Monetary Policy Surprise Spillovers," *Journal of International Economics*, vol. 75, pp. 180–96.

Duffee, G.R. (2013). "Forecasting Interest Rates," in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2B, Elsevier Publications, pp. 385–426.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, New York: 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.S. (2015). "Bond Risk Premia and Gaussian Term Structure Models." Available at SSRN 2320123.

Gagnon, J., Raskin, M., Remache, J. and Sack, B. (2011a). "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., Raskin, M., Remache, J. and Sack, B. (2011b). "Large-Scale Asset Purchases by the Federal Reserve: Did They Work?" Federal Reserve Bank of New York *Economic Policy Review,* vol. 17, pp. 41–59.

Gagnon, Joseph E. 2016. "Quantitative Easing: An Underappreciated Success," No. PB16-4.

Gambacorta, L., Hofmann, B. and Peersman, G. (2012). "The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis," Bank for International Settlements Working Paper.

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.

Goyal, A. and Welch, I. (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, vol. 21, pp. 1455–508.

Hamilton, J. and Wu, J. (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.

Joslin, S., Le, A. and Singleton, K.J. (2013). "Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs." *Journal of Financial Economics*, vol. 109(3), pp 604-622.

Joyce, M.A., Lasaosa, 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., Miles, D., Scott, A. and Vayanos, D. (2012). "Quantitative easing and unconventional monetary policy–an introduction," *The Economic Journal*, vol. 122(564), pp. F271-F288.

Kilian, L. (1998). “Small-Sample Confidence Intervals for Impulse Response Functions,” *Review of Economics and Statistics*, vol. 80, pp. 218–30.

Kirby, C. (1998). “The Restrictions on Predictability Implied by Rational Asset Pricing Models,” *Review of Financial Studies*, vol. 11, pp. 343–82.

Levich, R.M. and Potì, V. (2015). “Predictability and ‘Good Deals’ in Currency Markets,” *International Journal of Forecasting*, vol. 31(2), 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.

Mankiw, N.G. and Shapiro, M. (1986). “Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models.” *Economics Letters*, vol. 20, pp. 139–45.

Mark, N. (1995). “Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability,” *American Economic Review*, vol. 85, pp. 201–18.

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(4), 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.

Potì, V. and Siddique, A. (2013). “What Drives Currency Predictability?” *Journal of International Money and Finance*, vol. 36, pp. 86-106.

Potì, V. and Wang, D. (2010). “The Coskewness Puzzle,” *Journal of Banking and Finance*, vol. 34(8), pp. 1827–38.

Rapach, D.E. and Wohar, M.E. (2006). “In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining,” *Journal of Empirical Finance*, vol. 13(2), 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.

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.

Rudebusch, G.D. (1998a). "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review*, vol. 39(4), pp. 907–31.

Rudebusch, G.D. (1998b). "Do Measures of Monetary Policy in a VAR Make Sense? A Reply to Christopher A. Sims," *International Economic Review*, vol. 39(4), pp. 943–48.

Shapiro, M.D. and Watson, M. (1988). "Sources of Business Cycle Fluctuations," in Stanley Fischer, ed., *NBER Macroeconomics Annual 1988*, vol. 3, Cambridge, MA: MIT Press, pp. 111–48.

Sims, C. (1986). "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis *Quarterly Review*, Winter, pp. 2–16.

Sims, C.A. (1998). "Comment on Glenn Rudebusch's 'Do Measures of Monetary Policy in a VAR Make Sense?'" *International Economic Review*, vol. 39(4), pp. 933–41.

Stambaugh, R. (1986). "Bias in Regressions with Lagged Stochastic Regressors," Center for Research in Security Prices Working Paper, University of Chicago.

Stock, J.H. and Watson, M.W. (2003). "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, vol. 41(3), pp. 788–829.

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.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58(1), pp. 267–88.

Wang, H., Li, G. and Jiang, G. (2007). "Robust Regression Shrinkage and Consistent Variable Selection through the LAD-Lasso," *Journal of Business & Economic Statistics*, vol. 25(3), pp. 347–55.

Wang, H., Zhang, X., & Zou, G. (2009). Frequentist model averaging estimation: a review. Journal of Systems Science and Complexity, 22(4), 732-748.

Woodhill, L. (2013). "Bernanke's Blind QE Worship Amounts To Faith-Based Economics," *Forbes*, February 27, 2013; http://www.forbes.com/sites/louiswoodhill/2013/02/27/bernankes-blind-qe-worship-amounts-to-faith-based-economics/

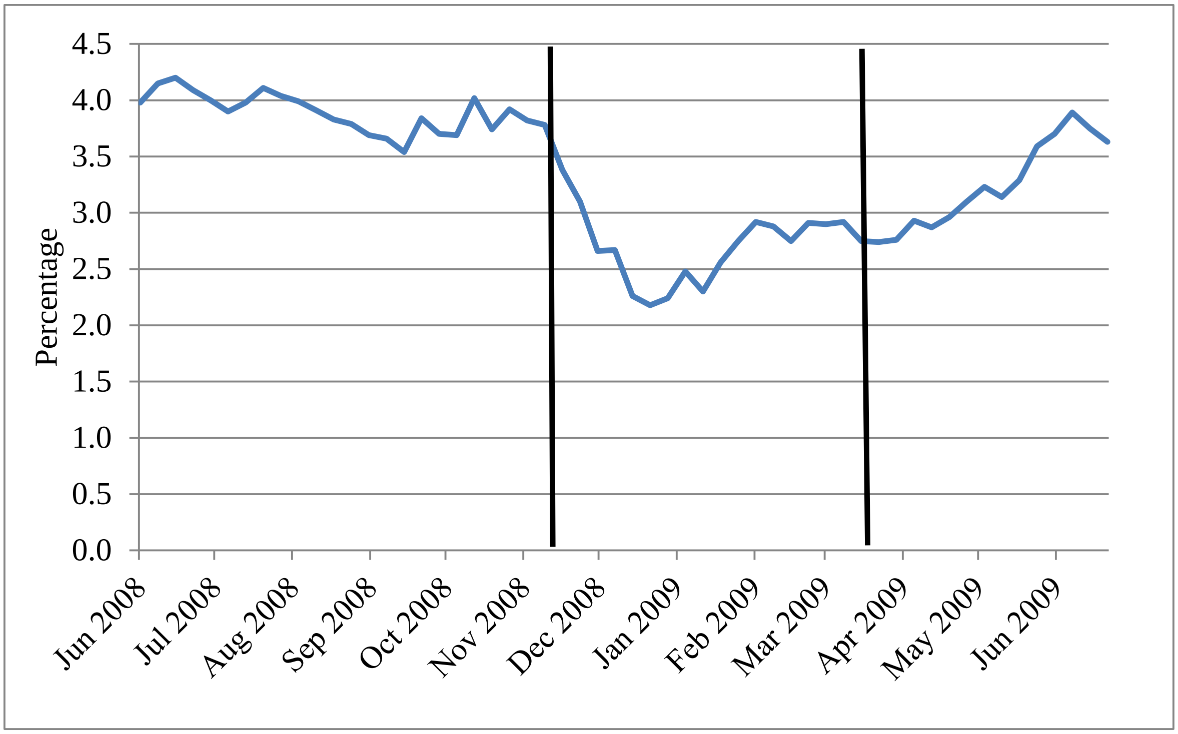
Wright, J.H. (2011). "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset." *American Economic Review*, vol. 101(4), pp. 1514–1534.

Wright, J.H. (2012). "What Does Monetary Policy Do to Long-Term Interest Rates at the Zero Lower Bound?" *The Economic Journal*, vol. 122, pp. 447–66.

Wright, J.H. (2014). “Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset: Reply,” *American Economic Review*, vol. 104(1), pp. 338–41.

Yu, E., (2016). “Did quantitative easing work?.” *Economic Insights*, vol. 1(1), pp.5-13.

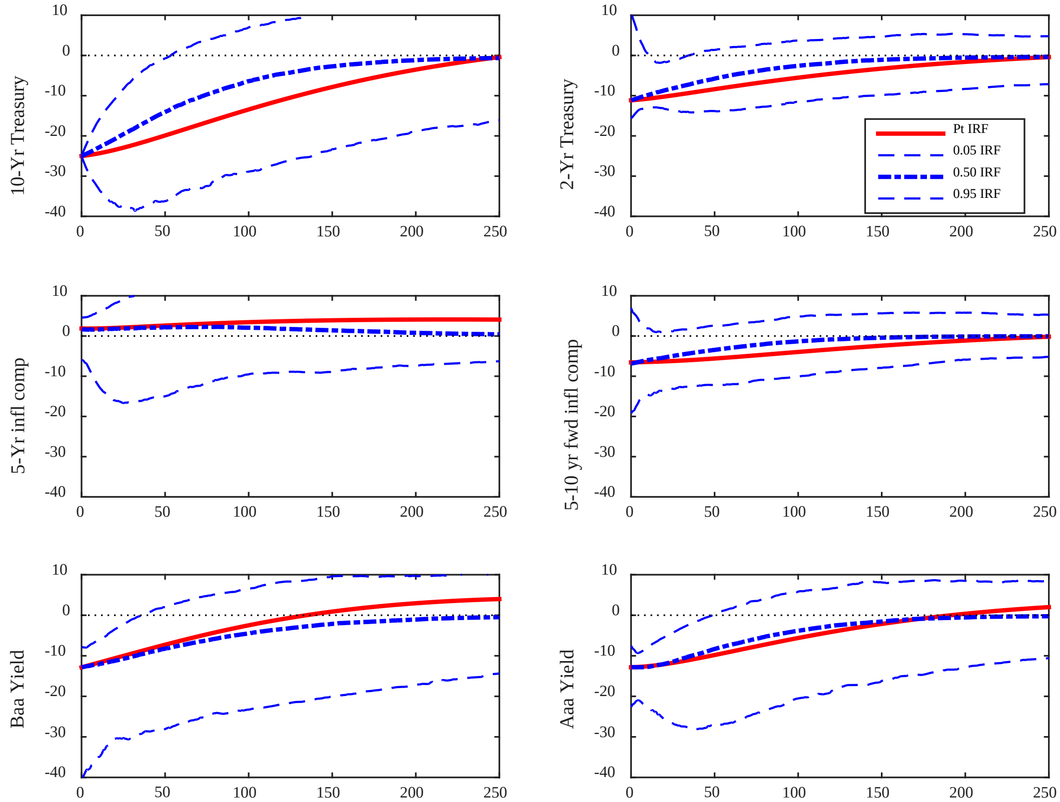
Figure 1: Nominal yields on 10-year Treasuries




Notes: The figure depicts yields on 10-year Treasuries from June 2008 through June 2009. Vertical lines denote the dates of LSAP announcements on November 25, 2008, and March 18, 2009.

Source: FRED®, Federal Reserve Economic Data, Federal Reserve Bank of St. Louis.

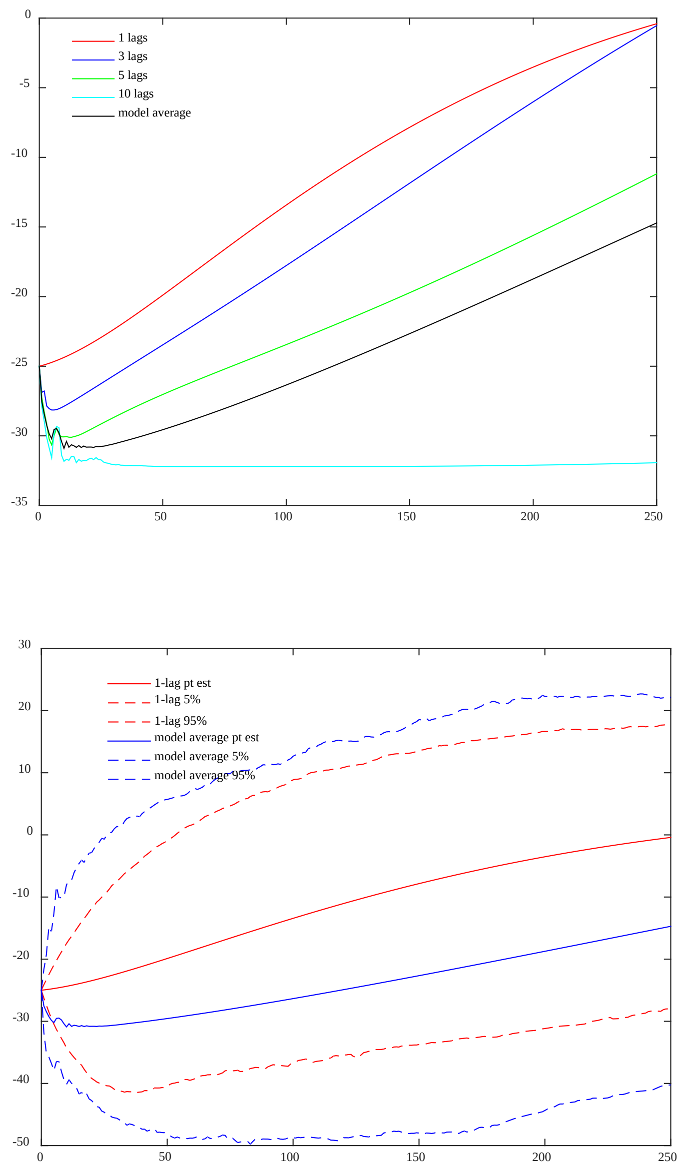
Figure 2: Impulse responses from the baseline VAR




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 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 unconventional monetary policy shocks. infl. comp., inflation compensation.

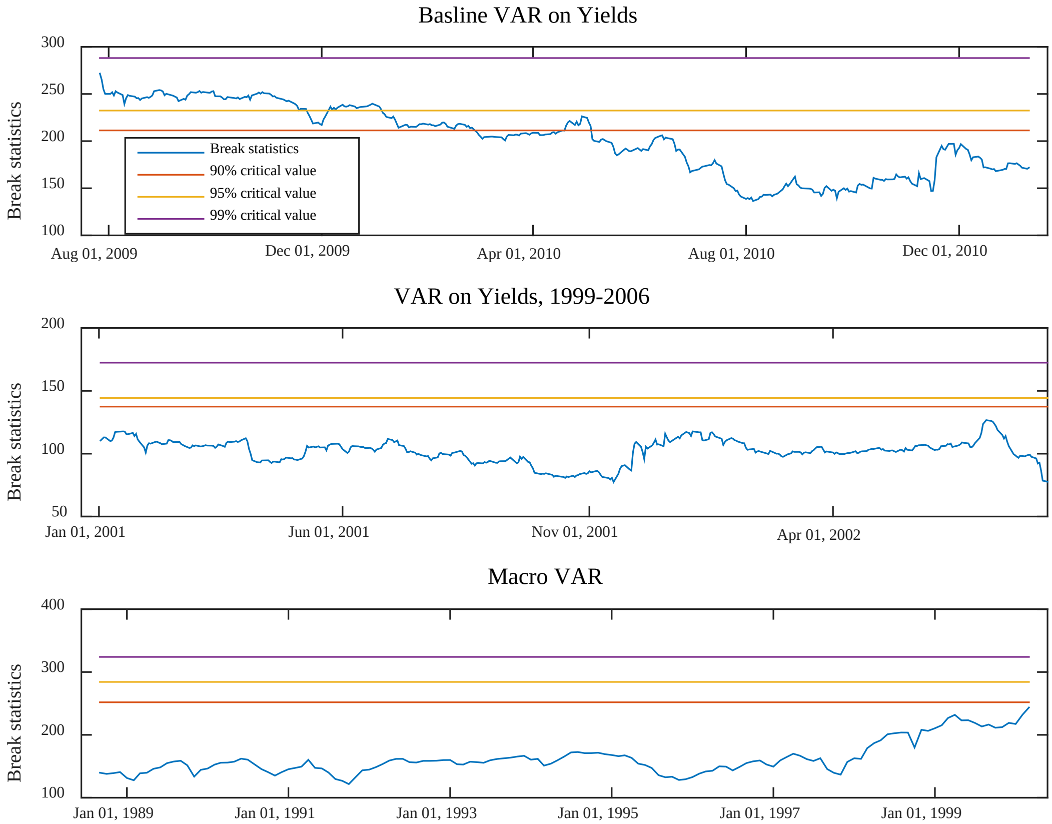
Source: Haver Analytics.

Figure 3: Impulse responses from the baseline VAR with alternative lag lengths



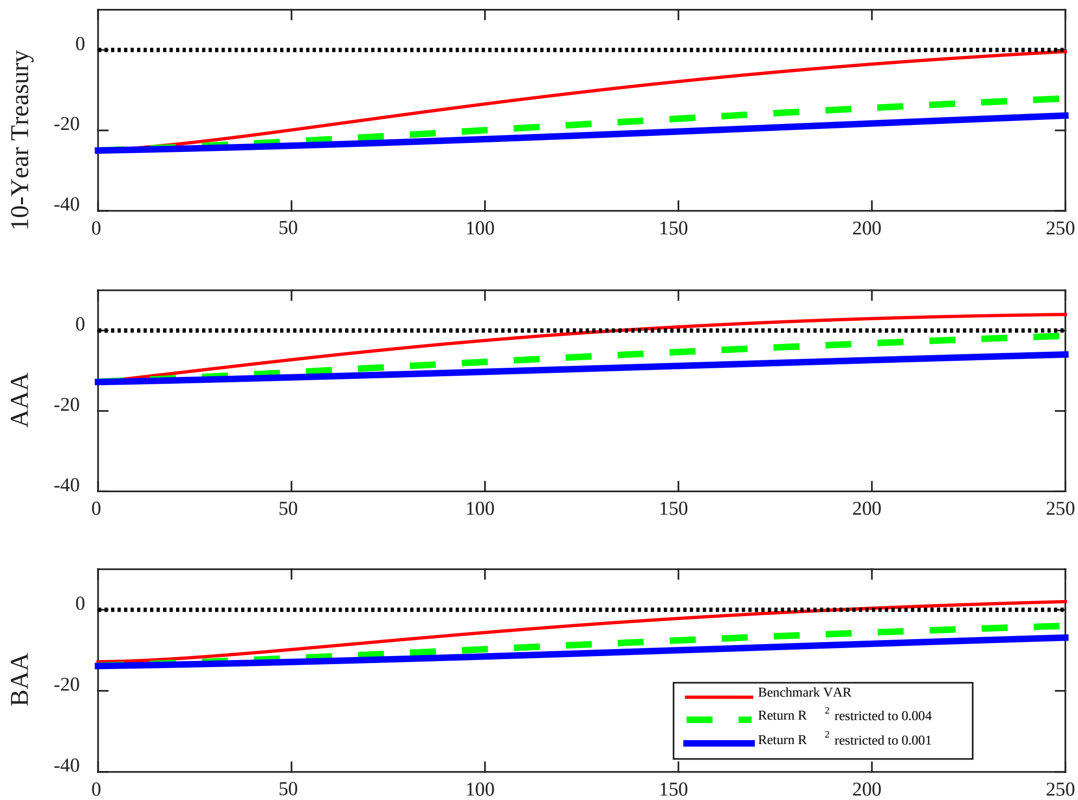

Notes: The top panel of the figure illustrates the responses in basis points of yields on the 10-year Treasury to unconventional monetary policy shocks using VARs with 1, 3, 5 and 10 lags, as well as a frequentist model-average of VARs with lags from 1 to 15. The bottom panel of the figure shows the 1-lag (red) and model averaged (blue) impulse responses to yields on the 10-year Treasury using VARs 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 4: Andrews (1993) structural break statistics for the VAR




Notes: The three panels of the figure plot the Andrews test statistics and the bootstrapped 1, 5, and 10 percent critical values from the 25$^{th}$ to the 75$^{th}$ percentile of the samples, for a structural break in three VARs: 1) the baseline VAR in yields, estimated from September 2008 through November 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 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, Aaa, and Baa 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—Treasuries 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-yr, 10-yr fwd infl comp | Baa Yield | Aaa Yield |
|---|---|---|---|---|---|---|
| 1-day Naive RMSFE | 0.02 | 0.05 | 0.03 | 0.04 | 0.05 | 0.05 |
| 20-day Naive RMSFE | 0.06 | 0.22 | 0.15 | 0.13 | 0.19 | 0.18 |
| 60-day Naive RMSFE | 0.07 | 0.40 | 0.30 | 0.18 | 0.31 | 0.32 |
| 120-day Naive 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 | 0.18 | 0.91 | 0.96 | 0.77 | 0.50 |
| 20-day Theil *p*-value | 0.02 | 0.12 | 0.89 | 0.74 | 0.69 | 0.43 |
| 60-day Theil *p*-value | 0.04 | 0.21 | 0.85 | 0.56 | 0.68 | 0.46 |
| 120-day Theil *p*-value | 0.17 | 0.32 | 0.88 | 0.56 | 0.70 | 0.54 |

Notes: This table shows OOS root mean square forecast statistics from the 6-variable VAR. The top panel shows the RMSFE for the naive (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 bootstrapped 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. fwd, forward; infl comp, inflation compensation.

Table 2: Out-of-sample mean error (bias) forecast statistics, constructed in an ex post sample

| | 2-Yr Treasury | 10-yr Treasury | 5-yr infl comp | 5-yr, 10-yr fwd infl comp | Baa Yield | Aaa Yield |
|---|---|---|---|---|---|---|
| 1-day Naive ME | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 20-day Naive ME | 0.00 | 0.02 | 0.01 | 0.01 | 0.00 | 0.02 |
| 60-day Naive ME | 0.01 | 0.09 | 0.03 | 0.02 | 0.02 | 0.08 |
| 120-day Naive ME | 0.03 | 0.18 | 0.01 | 0.02 | 0.04 | 0.17 |
| | | | | | | |
| 1-day Naive ME *t*-statistics | 0.05 | 0.77 | 0.58 | 0.47 | 0.22 | 0.71 |
| 20-day Naive ME *t*-statistics | 0.05 | 0.62 | 0.44 | 0.29 | 0.00 | 0.91 |
| 60-day Naive ME *t*-statistics | 0.75 | 0.81 | 0.32 | 0.45 | 0.19 | 0.95 |
| 120-day Naive ME *t*-statistics | 0.99 | 0.82 | 0.10 | 0.29 | 0.20 | 0.86 |
| | | | | | | |
| 1-day VAR ME | –0.01 | –0.02 | 0.00 | 0.00 | –0.01 | –0.02 |
| 20-day VAR ME | –0.13 | –0.38 | –0.04 | –0.07 | –0.19 | –0.32 |
| 60-day VAR ME | –0.19 | –0.75 | –0.01 | –0.17 | –0.42 | –0.62 |
| 120-day VAR ME | –0.22 | –0.95 | 0.03 | –0.22 | –0.57 | –0.77 |
| | | | | | | |
| 1-day VAR ME *t*-statistics | –11.85 | –8.89 | –2.89 | –1.05 | –4.32 | –8.84 |
| 20-day VAR ME *t*-statistics | –7.35 | –6.78 | –1.70 | –3.63 | –5.58 | –6.99 |
| 60-day VAR ME *t*-statistics | –5.60 | –4.59 | –0.07 | –4.49 | –4.28 | –4.61 |
| 120-day VAR ME *t*-statistics | –5.03 | –3.81 | 0.28 | –4.75 | –3.28 | –3.52 |

Notes: This table shows OOS mean error (ME) forecast statistics from the 6-variable VAR. The top panel shows the ME for the naive (martingale) forecast; the second panel shows the *t*-statistics for those naive mean errors, constructed with Newey-West statistics using appropriate lag length for the forecast horizon; the third and fourth panels show the same statistics for the bias-adjusted VAR coefficients. The OOS period is from October 2011 through November 2013.

Table 3: Bai-Perron structural stability tests

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

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 4: 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 | –14.7 | 0.1 | 0.4 |
| 10-Yr Treasury return | 2.0 | –6.4 | 0.1 | 0.4 |
| 5-Yr Δ inflation comp | 1.1 | 0.6 | 0.1 | 0.4 |
| 5-Yr-fwd Δ inflation comp | 2.1 | 1.1 | 0.1 | 0.4 |
| Baa return | 2.2 | –3.4 | 0.1 | 0.4 |
| Aaa return | 2.7 | –7.1 | 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 an annualized standard deviation to the market return of 20 percent. 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 5: Out-of-sample root mean square forecast statistics for the restricted VAR

| | 2-yr Treasury | 10-yr Treasury | 5-yr infl comp | 5-yr, 10-yr fwd infl comp | Baa Yield | Aaa Yield |
|---|---|---|---|---|---|---|
| Bias-corrected VAR | | | | | | |
| 1-day Theil | 1.16 | 1.09 | 1.00 | 0.99 | 1.02 | 1.12 |
| 20-day Theil | 1.94 | 1.92 | 0.99 | 0.94 | 1.51 | 2.32 |
| 60-day Theil | 2.13 | 1.86 | 1.11 | 1.03 | 1.84 | 2.22 |
| 120-day Theil | **2.09** | 1.72 | 1.02 | 0.99 | 1.91 | 1.82 |
| | | | | | | |
| $R^2$ restricted to 0.004 | | | | | | |
| 1-day Theil | 1.05 | 1.03 | 0.99 | 0.99 | 1.01 | 1.03 |
| 20-day Theil | 1.64 | 1.52 | **0.96** | **0.97** | 1.27 | 1.62 |
| 60-day Theil | 2.48 | 1.74 | **0.94** | 0.97 | 1.54 | 1.88 |
| 120-day Theil | 2.67 | 1.76 | **0.79** | 0.91 | 1.62 | 1.71 |
| | | | | | | |
| $R^2$ restricted to 0.001 | | | | | | |
| 1-day Theil | **1.01** | **1.01** | **0.99** | **0.99** | **1.00** | **1.00** |
| 20-day Theil | **1.23** | **1.16** | 0.99 | 0.97 | **1.08** | **1.18** |
| 60-day Theil | **1.84** | **1.30** | 1.02 | **0.93** | **1.20** | **1.35** |
| 120-day Theil | 2.24 | **1.39** | 0.86 | **0.82** | **1.28** | **1.35** |

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.