ifo Working Papers

Forecast Accuracy of a BVAR under
Alternative Specifications of the Zero Lower Bound

Tim Oliver Berg

Update March 2016

Forecast Accuracy of a BVAR under Alternative Specifications of the Zero Lower Bound

Update March 2016

Abstract

This paper discusses how the forecast accuracy of a Bayesian vector autoregression (BVAR) is affected by introducing the zero lower bound on the federal funds rate. As a benchmark I adopt a common BVAR specification, including 18 variables, estimated shrinkage, and no nonlinearity. Then I entertain alternative specifications of the zero lower bound. I account for the possibility that the effect of monetary policy on the economy is different in this regime, replace the federal funds rate by its shadow rate, consider a logarithmic transformation, feed in monetary policy shocks, or utilize conditional forecasts allowing for all shocks implemented through a rejection sampler. The latter two are also coupled with interest rate expectations from future contracts. It is shown that the predictive densities of all these specifications are greatly different, suggesting that this modeling choice is not innocuous. The comparison is based on the accuracy of point and density forecasts of major U.S. macroeconomic series during the period 2009:1 to 2014:4. The introduction of the zero lower bound is not beneficial *per se*, but it depends on how it is done and which series is forecasted. With caution, I recommend the shadow rate specification and the rejection sampler combined with interest rate expectations to deal with the nonlinearity in the policy rate. Since the policy rate will remain low for some time, these findings could prove useful for practical forecasters.

Tim Oliver Berg
Ifo Institute – Leibniz Institute for
Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1689
berg@ifo.de

1 Introduction

Bayesian vector autoregressions (BVARs) are useful tools to forecast major macroeconomic time series. In contrast to classical VARs, they do not entail the risk of overparameterization and are hence able to exhaust the information contained in large datasets. A popular approach is that of Bańbura, Giannone, and Reichlin (2010), who use a variant of the Minnesota prior as proposed by Doan, Litterman, and Sims (1984) and Litterman (1980, 1986). Their prior shrinks forecasts towards a random walk for those variables showing substantial persistence and white noise for those that do not. The degree of shrinkage is controlled by a single hyperparameter, which can be easily computed or estimated in an optimal fashion.

In a forecast experiment the researcher has to make lots of choices with respect to model specification, estimation, and forecast construction. Because the specification choices affect the forecast accuracy of a BVAR, the literature has discussed issues such as: the appropriate number of variables included, whether variables should enter in (log-)levels or growth rates, whether a (log-)level specification should be complemented with priors favoring unit roots and cointegration, the treatment of the residual covariance matrix, the lag length, the selection of the shrinkage parameter, whether a recursive or rolling window estimation scheme should be used, and whether forecasts should be computed by iteration or the direct-step approach (see, e.g., Koop, 2013; Berg and Henzel, 2015; Carriero, Clark, and Marcellino, 2015, among others).

This paper discusses a further specification choice, which has not been addressed carefully in the literature. In particular, I investigate how the forecast accuracy of a BVAR is affected by introducing a zero lower bound on the federal funds rate. Researchers have dealt with the possibility that the monetary policy instrument is constraint from below by either ignoring it or by adjusting their evaluation period such that the recent era of exceptionally low interest rates is avoided. For instance Clark and McCracken (2014) write: "We stop our forecast evaluation in 2007:Q4 to avoid possible complications of the zero lower bound constraints that became relevant in subsequent years." (see p. 27, ll. 25-26).[1]

What Clark and McCracken (2014) mean by the relevance of the zero lower bound constraint becomes obvious when inspecting Figure 1. The figure shows the federal funds rate and forecasts thereof one quarter and four quarters ahead. While the federal funds rate is stuck between

---

[1] I admit that there are no technical or legal reasons that constrain the policy rate at zero. In fact, some countries have already experimented with (slightly) negative deposit or repo rates. However, such rates cannot go that far below zero as long as paper currency is available (see, e.g., Rogoff, 2015). The BVAR specifications presented in this paper are able to include this feature. The lower bound could be reduced to equal the cost of cash hoarding, say to -0.5% or -1%, and forecasts below that cost would receive zero probability. In this paper I do not follow that avenue, but take the zero lower bound constraint literally.
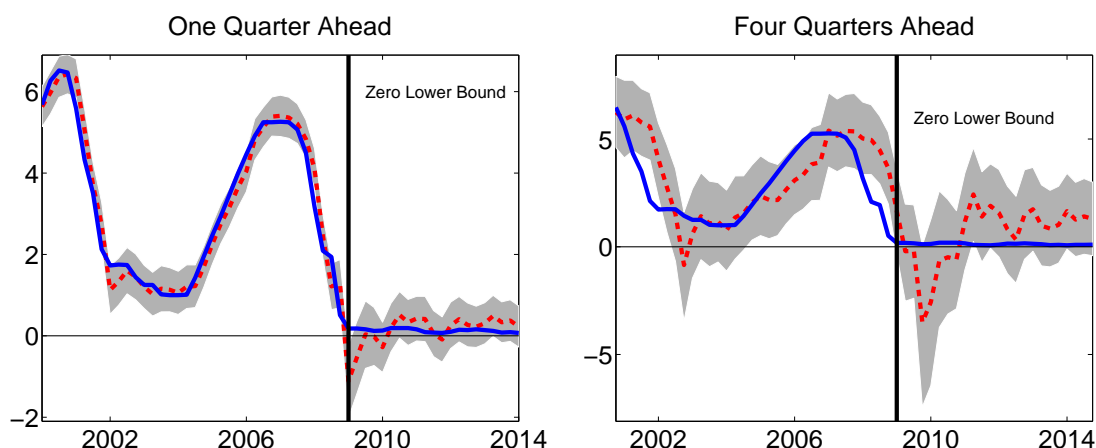
Figure 1: Federal Funds Rate – Forecast and Realization. Notes: this figure shows the point forecast (dashed line) together with the realization (solid line) and a 90% error band (shaded area) for the periods 2000:1 to 2014:1 (left panel) and 2000:4 to 2014:4 (right panel). x-axis: year; y-axis: percent.

0% and 0.25% since 2009, the BVAR attaches a lot probability mass to negative policy rates. And the BVAR also suggests a federal funds rate liftoff as early as mid-2011, despite the Federal Reserve's repeated promise to keep interest rates low for an "extended period." Such statements were part of the central bank's policy to anchor longer-term interest rate expectations during the zero lower bound period, known as forward guidance (see Bernanke, 2013).

The zero lower bound constraint and the Federal Reserve's forward guidance are challenges for any forecast model. However, ignoring them or discarding years of observations are no good options. And since the federal funds rate will most likely remain low for some time even after a liftoff, the results in this paper could prove useful for practical forecasters and researchers alike.

Hence I focus on the period from 2009 to 2014 and entertain alternative specifications of the zero lower bound. Also, I include interest rate expectations from future contracts to account for the Federal Reserve's forward guidance. Accordingly, most BVAR specifications considered in this paper are not linear and forecasts have nonstandard distributions.[2]

As a benchmark I adopt a common BVAR specification, including 18 variables that enter in (log-)levels, while the priors imposed favor unit roots and cointegration. Moreover, I use a long lag specification, while parsimony is achieved by optimal estimation of the shrinkage parameter as proposed by Giannone, Lenza, and Primiceri (2015). Forecasts are obtained by iteration using a recursive estimation scheme. My reading of the literature is that such a specification produces relatively accurate point and density forecasts for major macroeconomic series in normal times.

---

[2]For linear models with Gaussian shocks, predictive densities are approximately multivariate normal (see Gelman, Hwang, and Vehtari, 2014).

Then I entertain alternative specifications of the zero lower bound. First, I account for the possibility that the effect of monetary policy on the economy is different below a threshold. Second, I replace the federal funds rate by the shadow rate of Wu and Xia (2015). Third, I substitute the federal funds rate by its natural logarithm. Fourth, I feed shocks to monetary policy into the BVAR to prevent the policy rate from falling below zero. Fifth, I use the soft condition approach of Waggoner and Zha (1999) to truncate the predictive density of the federal funds rate at zero. Finally, I couple the latter two specifications with interest rate expectations from future contracts to rule out federal funds rate predictions that appear too high. As a first result I obtain that the alternative specifications produce very different predictive densities for the federal funds rate.

In the next step I discuss how this finding impacts on the forecast accuracy of the BVAR with respect to other key variables. Particularly, I rank the alternative specifications according to their point and density forecast accuracy for GDP growth, the unemployment rate, CPI inflation, and the treasury rate during the period from 2009:1 to 2014:1 up to four quarters ahead. For a robust assessment I consider several relative and absolute forecast metrics.

I obtain a large set of empirical results that can be summarized as follows. First, introducing the zero lower bound hardly affects the forecast accuracy of the BVAR for GDP growth, CPI inflation, and the unemployment rate one quarter ahead but does four quarters ahead. While this is by construction for the policy shock specifications, it is also observed for the other specifications. A practical forecaster should thus be more concerned about the zero lower bound when she is interested in long-term forecasts. Even for the treasury rate this specification choice does not matter in the short-term. Second, the performance of the alternative specifications differs four quarters ahead, suggesting that this modeling choice is not innocuous. In particular, the introduction of the zero lower bound is not beneficial *per se*, but it depends on how it is done and which series is forecasted.

With caution, I recommend the shadow rate specification. Except for GDP growth, this specification delivers a satisfactory forecast ability in absolute terms and relative to the benchmark. In contrast, the threshold and logarithmic specifications are no good choices, while both policy shock specifications are as accurate as the benchmark. Finally, the soft condition specification is useful only in combination with external information, but then helps to achieve a better calibration of predictive densities and reduces forecast errors relative to the benchmark for all series, except the unemployment rate.

The paper is organized as follows. Section 2 discusses the BVAR and specification choices. Section 3 presents the alternative specifications of the zero lower bound. Section 4 describes the data and presents the forecast comparison. Section 5 summarizes the results and concludes.

2   Bayesian Vector Autoregression

Consider the following VAR

$$y_t = c + B_1 y_{t-1} + ... + B_p y_{t-p} + u_t, \tag{1}$$

where y_t is a $n \times 1$ vector of endogenous variables; c is a $n \times 1$ vector of intercepts; B_i are $n \times n$ matrices of coefficients; $i = 1, ..., p$ denotes the lags included; u_t is a $n \times 1$ vector of normally distributed residual terms with zero mean and covariance matrix Σ; and data are available for $t = 1 - p, ..., T$. Finally, denote the total number of coefficients by $k = n(1 + np)$.

In the forecast experiment the VAR is estimated on $n = 18$ variables including $p = 4$ lags of each (hence $k = 1,314$).[3] Such a large dimensional system of multivariate regressions is, however, not estimable without imposing additional prior beliefs on the parameters. Hence I follow common practice and use a variant of the Minnesota prior to deal with the dense parameterization of the model. The basic idea is that a random walk with drift is a reasonable description of the data generating process behind most macroeconomic series when considered in (log-)levels. For data in differences a white noise process is the appropriate prior. The VAR is hence centered around the prior mean:

$$y_t = c + \delta y_{t-1} + u_t, \tag{2}$$

and imposing the prior amounts to shrinking most coefficients towards zero. The hyperparameter δ is set equal to 1 for the levels specification and to 0 for the difference specification.

In contrast to the original Minnesota prior (see Litterman, 1980, 1986), I do not assume Σ to be known and diagonal, but use the generalized version of Kadiyala and Karlsson (1993, 1997) that allows for correlation among residuals. The evidence in Bańbura et al. (2010) suggests that a generalized Minnesota prior produces accurate forecasts for major macroeconomic series even though the $n(n+1)/2$ distinct elements of Σ have to be estimated on top of the k coefficients.

To further improve the forecast accuracy of BVARs, the literature proposes to complement the Minnesota prior with additional priors that favor unit roots and cointegration, and introduce correlation among the coefficients in a given equation. The first prior is the "sum-of-coefficients" prior by Doan et al. (1984), while the second prior is proposed by Sims (1993) and known as the "dummy-initial-observation" prior. Both are implemented by augmenting the VAR in (1) with dummy observations. The implementation and estimation details are provided in Appendix A.

---

[3]The choice for the lag length can be motivated by the frequency of the data, which is quarterly.

3 Alternative Specifications of the Zero Lower Bound

In this section I present alternative specifications that introduce the zero lower bound into the BVAR. Furthermore, I discuss how the predictive density of the federal funds rate is affected.

3.1 Specifications

In particular, I consider the following specifications:

Benchmark In the first specification I neglect the presence of the zero lower bound and construct forecasts using a standard BVAR. This procedure serves as the benchmark for all the other specifications. While the federal funds rate is constrained to be nonnegative, forecasts thereof may take negative values. And with the federal funds rate stuck between 0% and 0.25% since 2009, the policy instrument may progressively lose importance for the U.S. economy over time.

Threshold In the second specification I still allow for negative federal funds rate forecasts, but account for the possibility that the effect of monetary policy on the economy is different in a low policy rate regime. In particular, I consider a two-regime threshold BVAR of the form:[4]

$$y_t = c_1 + B_1\left(L\right)y_t + u_{1,t} + \left(c_2 + B_2\left(L\right)y_t + u_{2,t}\right)\mathbb{I}\left[R_{t-d} < \theta\right], \tag{3}$$

where L denotes the lag operator, and $\mathbb{I}$ is an indicator function that is equal to 1 whenever the federal funds rate (R_t) is below a certain threshold (θ) at time $t-d$ and 0 otherwise. The approach hence allows for both regime dependent coefficients and heteroskedasticity. Moreover, I fix the threshold value at the first quartile of the historical federal funds rate distribution ($\theta = 3.195\%$). On the one hand, the threshold should be relatively low such that the regime primarily captures the dynamics in the proximity of the zero lower bound. On the other hand, the number of time series observations may not become too small. The value chosen seems to account well for both concerns. Finally, I set $d = 0$ and estimate the threshold BVAR for each regime separately using the benchmark routine.[5]

Shadow Rate In the third specification I replace the federal funds rate from 2009:1 onwards by the shadow short-term interest rate of Wu and Xia (2015)[6] and then construct forecasts as usual. Unlike the federal funds rate, the shadow rate can fall below zero in unconventional monetary

---

[4]See Hubrich and Teräsvirta (2013) for a survey on the specification and estimation of threshold VARs.

[5]The choice for the delay parameter d and the threshold θ has no impact on the qualitative results.

[6]The Wu-Xia data are available here: https://www.frbatlanta.org/cqer/research/shadow_rate.aspx. I also experimented with the shadow rate estimates by Krippner (2013) with similar qualitative results.

policy environments (see also Figure B.1). Intuitively, it is the nominal interest rate with shortest maturity that would prevail in the absence of physical currency. Wu and Xia (2015) derive their shadow rate from a nonlinear term structure model estimated on forward rates and show that it interacts with the U.S. economy since 2009 in the same fashion as the federal funds rate did before. The shadow rate also reflects the unconventional monetary policy measures implemented by the Federal Reserve in recent times and is hence a reasonable proxy for the monetary policy stance during the zero lower bound period. Still, the hybrid federal funds/shadow rate series is hypothetical since economic agents cannot transact at shadow rates.

Logarithmic In the fourth specification I replace the federal funds rate by its natural logarithm. While the federal funds rate is bounded between 0 and ∞, the logarithmic transformation maps this into $(-\infty, \infty)$. The BVAR is hence estimated including $\log(R_t)$, and I obtain the predictive density for R_t using the inverse log transformation, which makes sure that all posterior draws for the federal funds rate are nonnegative. Cogley and Sargent (2002) follow a similar strategy to deal with the boundedness of the unemployment rate. Even though the logarithmic transformation truncates the predictive density of the federal funds rate at zero, it also introduces a nonlinear relation between the monetary policy instrument and the economy, which may not be supported by the data.[7]

Policy Shock In the fifth specification I follow the standard procedure of the DSGE forecasting literature and feed shocks to monetary policy into the BVAR to prevent the federal funds rate from falling below zero (see, e.g., Del Negro and Schorfheide, 2013, among others). Under this specification the ordering of the variables in y_t hence matters. In particular, I divide the sample into n_1 slow (S_t) and n_2 fast moving (F_t) variables and estimate the BVAR on $y_t = (S_t \ R_t \ F_t)'$. The former category includes production, employment, and price data, while the latter contains financial market variables. The monetary policy shock is then the innovation to R_t. Hitting the zero lower bound in $T + h$ hence alters forecasts for slow moving variables in $T + h + 1$ and later, while forecasts for fast moving variables are also affected in $T + h$. But none variable is affected in $T + h - 1$ and earlier since the monetary policy shocks are not anticipated.[8]

Soft Condition In the sixth specification I follow Waggoner and Zha (1999), incorporating the zero lower bound using an acceptance-rejection algorithm. In particular, parameter and shock

---

[7]In the logarithmic BVAR a decline of the federal funds rate from 10% to 5% has the same effect on the economy as a decline from 1% to 0.5%, which is hardly consonant with standard economic theory.

[8]The DSGE forecasting literature also considers anticipated monetary policy shocks since reaching the zero lower bound may not come as a surprise to economic agents. However, this is not possible in a VAR.

constellations that are not consistent with a nonnegative federal funds rate forecast receive zero prior weight and are discarded.[9] The Waggoner and Zha (1999) method can be computationally intense since the acceptance rate may be low in some periods. Moreover, this specification affects not only the predictive density, but also the posterior of the coefficients and the composition of the residual vector. For instance, shocks that imply a decrease in the policy rate, e.g. negative monetary policy shocks, receive a lower weight than those that lead to an increase, e.g. positive demand shocks.

External Information (EI) Finally, I combine the policy shock and soft condition specifications with external information, ruling out federal funds rate forecasts that appear too high. To obtain a reasonable upper bound, I use the expectations inherent in future contracts.[10] In particular, I restrict a federal funds rate forecast for the period $T + h$ to the interval $\hat{R}_{T+h}^j \in \left(0, 2 \times \hat{R}_{T+h}^f\right)$, where $\hat{R}_{T+h}^f$ denotes the financial market expectation. For example, if financial market participants expect the federal funds rate to be 0.5% next quarter, I ensure that forecasts from the BVAR are positive but not larger than 1%, which implies that the predictive mean roughly matches the market expectations. Conditioning forecasts on external information has become popular and proved to be beneficial (see, e.g., Clark and McCracken, 2014; Smets, Warne, and Wouters, 2014; Bańbura, Giannone, and Lenza, 2015, among others). In accordance with the procedures above, I implement the upper bound either by adding negative monetary policy shocks such that the federal funds rate equals the upper bound (policy shock specification) or by discarding draws that violate the restriction (soft condition specification).

3.2 Discussion

Before the different specifications are ranked according to forecast accuracy, I discuss how they affect the predictive densities of the federal funds rate, which are provided in Figure 2 for $h = 1$ and the period 2012:1 based on 25,000 posterior draws.[11] The figure suggests that the density is symmetric for the benchmark BVAR, while the dispersion is large. The benchmark also assigns a high probability to a negative federal funds rate. For the threshold BVAR the density is similar, but the dispersion is even larger as only a relatively small number of observations is available for

---

[9]This condition is soft in the sense that future values of the federal funds rate are restricted to a certain range, whereas a hard condition would fix forecasts at a single point, say at 1%.

[10]Federal funds rate future contracts are traded on the Chicago Board of Trade and historical prices are available from Datastream. For example, the mnemonic for the January 2012 contract is "CFF0112". See also Figure B.1 for a times series plot.

[11]I use this period as an example. The densities are qualitatively similar for other periods and horizons.
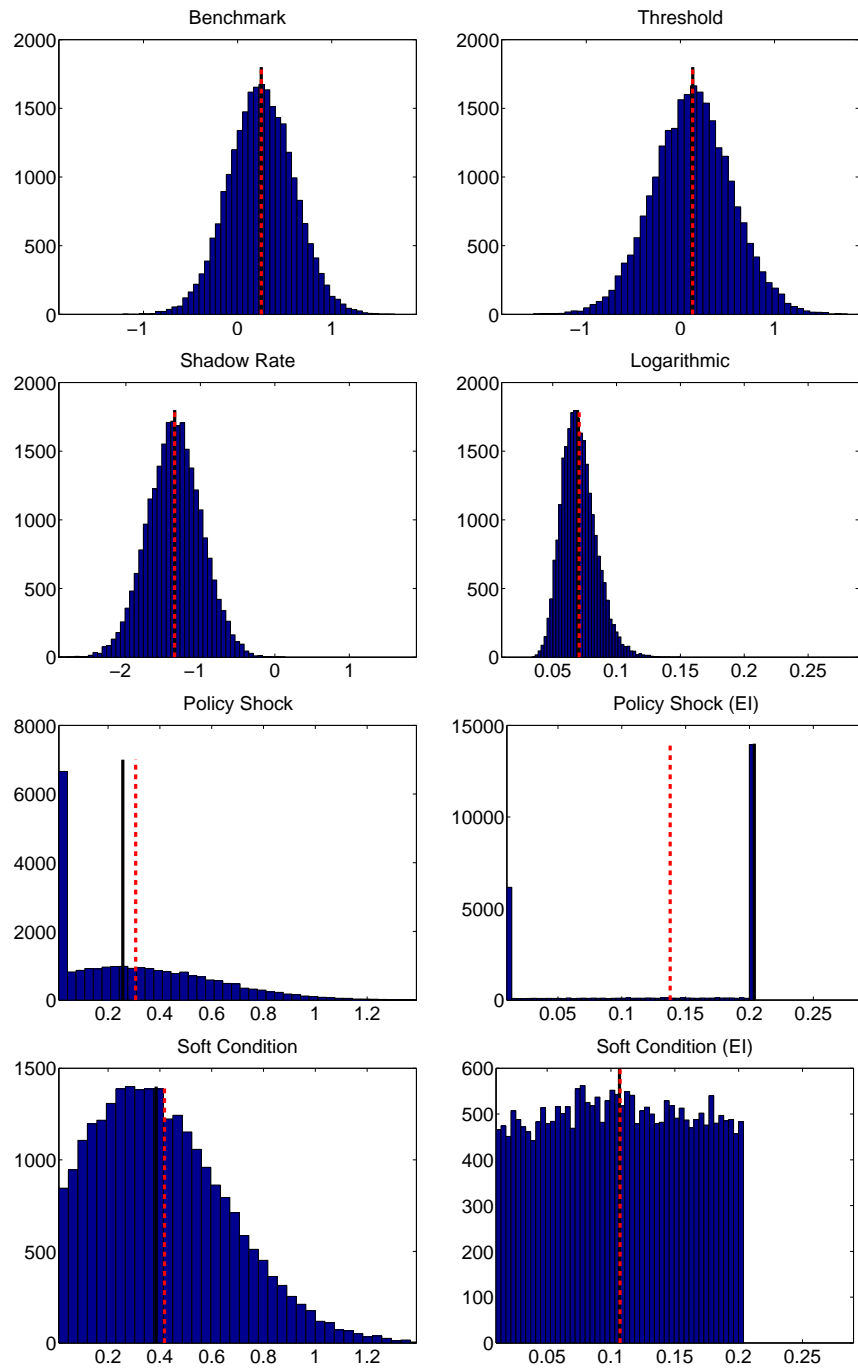
Figure 2: Predictive Density of Federal Funds Rate. Notes: this figure shows the predictive density of the federal funds rate (or its shadow rate) for $h = 1$ and the period 2012:1 under alternative specifications of the zero lower bound. The vertical lines denote the predictive mean (dashed) and median (solid), respectively. x-axis: percent; y-axis: number of posterior draws.

8

Table 1: Alternative Specifications of the Zero Lower Bound

Specification	Policy Instrument	Predictive Density of R_t (or S_t)			
		Density	Support	Symmetry	Skewness
Benchmark	R_t	Normal	$(-\infty, \infty)$	Yes	None
Threshold	R_t	Normal	$(-\infty, \infty)$	Yes	None
Shadow Rate	S_t	Normal	$(-\infty, \infty)$	Yes	None
Logarithmic	$\log(R_t)$	Log-Normal	$(0, \infty)$	No	Right
Policy Shock	R_t	n/a	$(0, \infty)$	No	Right
Policy Shock (EI)	R_t	n/a	$(0, b)$	No	Right/Left
Soft Condition	R_t	n/a	$(0, \infty)$	No	Right
Soft Condition (EI)	R_t	n/a	$(0, b)$	No	Right/Left

Notes: this table summarizes the main properties of the predictive density of the federal funds rate (R_t) or its shadow rate (S_t) under alternative specifications of the zero lower bound. n/a means that the density is nonstandard, while b denotes an upper bound derived from external information (EI).

the low interest rate regime, which increases estimation uncertainty. Replacing the federal funds rate by its shadow rate does neither affect the shape nor the dispersion of the density, but shifts it completely into negative territory, which is consonant with the shadow rate concept. Under the first three specifications the BVAR is linear and predictive densities are approximately Gaussian. Simulating the BVAR with $\log(R_t)$ rather than R_t introduces a nonlinearity. The density is log-normal, right-skewed, and takes only nonnegative values. It is also much more condensed than those of the first three specifications.

While the shadow rate and logarithmic specification build on replacing or transforming the policy rate before estimation, the policy shock and soft condition specifications include the federal funds rate, but manipulate its predictive density. For instance, the policy shock specification generates a density that is that of the benchmark, but truncated at zero. All negative draws are shifted by adding positive monetary policy shocks, producing a density with a lot probability mass at zero. When combined with external information, all draws above the upper bound are also shifted, and the density may be bimodal. The density under the soft condition specification is also truncated at zero, but less right-skewed. In combination with external information this specification leads to a density that appears uniform with mean close to the market expectations.

In sum, the alternative specifications produce very different predictive densities for the federal funds rate, and I provide a summary of their main properties in Table 1. Below I assess how this finding impacts on the forecast accuracy of the BVAR with respect to other key variables.

Table 2: Dataset

No.	Mnemonic	Label	Category	Source	S/R/F	Code
01	GDPC1	Real Gross Domestic Product	National Accounts	FRED	S	1
02	PCECC96	Real Personal Consumption Expenditures	National Accounts	FRED	S	1
03	GPDIC1	Real Gross Private Domestic Investment	National Accounts	FRED	S	1
04	INDPRO	Industrial Production Index	Business Activity	FRED	S	1
05	TCU	Capacity Utilization: Total Industry	Business Activity	FRED	S	3
06	NAPM	ISM Manufacturing: PMI Composite Index	Business Activity	FRED	S	3
07	USUMCONSH	University of Michigan: Consumer Sentiment	Business Activity	TRDS	S	3
08	PAYEMS	All Employees: Total Nonfarm	Employment	FRED	S	1
09	UNRATE	Civilian Unemployment Rate	Employment	FRED	S	3
10	COMPRNFB	Real Compensation Per Hour: Nonfarm	Employment	FRED	S	1
11	PCECTPI	PCE: Chain-type Price Index	Prices	FRED	S	1
12	CPIAUCSL	CPI for All Urban Consumers: All Items	Prices	FRED	S	1
13	PPIACO	Producer Price Index: All Commodities	Prices	FRED	S	1
14	FEDFUNDS	Effective Federal Funds Rate	Policy Instrument	FRED	R	3
15	DGS10	10-Year Treasury Constant Maturity Rate	Financial Markets	FRED	F	3
16	BAA	Moody's Seasoned Baa Corporate Bond Yield	Financial Markets	FRED	F	3
17	S&PCOMP	S&P 500 Composite - Price Index	Financial Markets	TRDS	F	2
18	TWEXBPA	Real Trade Weighted U.S. Dollar Index: Broad	Financial Markets	FRED	F	2

Notes: Mnemonic, label, and category as used in database. Source: FRED = Federal Reserve Economic Database; TRDS = Thomson Reuters Datastream. S = Slow moving; R = Policy instrument; F = Fast moving. Transformation code: $1 = 400 \cdot \log(x_t)$; $2 = 100 \cdot \log(x_t)$; $3 = $ None (Levels specification); $1 = 400 \cdot \log(x_t/x_{t-1})$; $2 = 100 \cdot \log(x_t/x_{t-1})$; $3 = $ None (Difference specification).

4 Forecast Comparison

In this section I first present the data and explain the forecast experiment. Thereafter I compare the alternative specifications in terms of point and density forecast accuracy measures.

4.1 Data and Forecast Experiment

The dataset includes 18 quarterly U.S. macroeconomic series for the period 1985:1 to 2014:4. The starting point coincides roughly with the beginning of the Great Moderation era and is also chosen to avoid the monetarist experiment of the early 1980s. In most cases the series are obtained from the Federal Reserve Economic Database (FRED) and cover the following categories that are particularly relevant for all policymakers: national accounts, business activity, employment, prices, financial markets, and the federal funds rate as the monetary policy instrument.

Except for diffusion indexes from survey data (e.g., consumer sentiment) and those variables that are expressed in rates (e.g., the unemployment rate), I include all of them in log-levels. This proceeding is consistent with the random walk prior imposed on the VAR coefficients ($\delta = 1$).[12] I provide a detailed description of the series and their exact transformation in Table 2. The table also shows the ordering of the variables, which is relevant for the policy shock specification.

---

[12]The results for the difference specification ($\delta = 0$) are reported in Appendix B.

Table 3: Forecast Metrics

Metric	Short	Reference	Forecast	Evaluation	Measure
Mean Forecast Error	MFE	Point	Mean	Absolute	pp
Root Mean Squared Forecast Error	RMSFE	Point	Mean	Relative	pp
Mean Absolute Forecast Error	MAFE	Point	Mean	Relative	pp
Log Predictive Density Score	LPDS	Density	PDF	Relative	n/a
Continuous Ranked Probability Score	CRPS	Density	CDF	Relative	pp
Probability Integral Transform	PIT	Density	CDF	Absolute	n/a

Notes: this table summarizes the main properties of the forecast accuracy measures. The acronyms mean: PDF = Probability density function; CDF = Cumulative density function; pp = Percentage points.

For each specification I produce forecasts up to $H = 4$. I begin with the estimation sample 1985:1 to 2008:4 and generate forecasts for 2009:1 to 2009:4. I then iterate forward, always updating the estimation sample one quarter at a time, until 2013:4, producing forecasts for 2014:1 to 2014:4.[13] This procedure yields a sequence of 21 forecast sets for each specification. The evaluation period hence runs from 2009:1 to 2014:1 for the shortest forecast horizon ($h = 1$) and from 2009:4 to 2014:4 for the longest ($h = 4$), and coincides with the recent zero lower bound period.

4.2 Forecast Metrics

The different specifications are compared in terms of several point and density forecast accuracy measures. Each metric represents an alternative loss function and computing all of them allows for a robust assessment of forecast performance. I summarize their main properties in Table 3.

For point forecasts I begin with the mean forecast error (MFE). Let $\bar{y}^j_{i,T|T-h}$ be the predictive mean for variable i produced by approach j and $y_{i,T}$ the realization. The MFE is computed as

$$\text{MFE}^j_{i,h} = \frac{1}{T_1 - T_0 + 1} \sum_{T=T_0+h}^{T_1+h} \left(\bar{y}^j_{i,T|T-h} - y_{i,T} \right),\tag{4}$$

where $T_0 = 96$ and $T_1 = 116$ denote the last observation of the first (2008:4) and final (2013:4) estimation sample, respectively. The MFE allows for an absolute assessment of the point forecast accuracy because it detects any systematic bias of the predictive mean. Whenever the MFE is positive (negative) the BVAR overpredicts (underpredicts) a particular variable.

[13]Berg and Henzel (2015) show that BVARs benefit a lot in terms of forecast accuracy from recursive estimation schemes that use all available observations up to period T compared to rolling window schemes that discard the most distant observations. Therefore I prefer the former to the latter.

11

For a forecaster with a quadratic loss function the predictive mean is also the optimal forecast and the root mean squared forecast error (RMSFE) is the appropriate measure to discriminate among forecast models (see Weiss, 1996):

$$\text{RMSFE}_{i,h}^{j} = \sqrt{\frac{1}{T_1 - T_0 + 1} \sum_{T=T_0+h}^{T_1+h} \left(\bar{y}_{i,T|T-h}^{j} - y_{i,T} \right)^2}. \tag{5}$$

Since the RMSFE is sensitive to large forecast errors, I also compute the mean absolute forecast error (MAFE), which puts less weight on extreme observations:[14]

$$\text{MAFE}_{i,h}^{j} = \frac{1}{T_1 - T_0 + 1} \sum_{T=T_0+h}^{T_1+h} |\bar{y}_{i,T|T-h}^{j} - y_{i,T}|. \tag{6}$$

While the MFE is a useful measure to evaluate the absolute accuracy of a point forecast, the RMSFE and MAFE are informative only when compared across models. Thus I will report them for each specification j relative to the benchmark B:

$$\left(\text{RMSFE}_{i,h}^{j}/\text{RMSFE}_{i,h}^{B} \right) \quad \text{and} \quad \left(\text{MAFE}_{i,h}^{j}/\text{MAFE}_{i,h}^{B} \right). \tag{7}$$

For density forecasts I start with the log predictive density score (LPDS), which is a standard metric to compare forecast models in a Bayesian setting (see, e.g., Geweke and Amisano, 2010, among others). The LPDS is obtained by computing the height of the predictive density at the realization, and I report its average across evaluation periods:

$$\overline{\text{LPDS}}_{i,h}^{j} = \frac{1}{T_1 - T_0 + 1} \sum_{T=T_0+h}^{T_1+h} \log p_{i,T|T-h}^{j} \left(y_{i,T} \right), \tag{8}$$

where $p_{i,T|T-h}^{j} \left(\cdot \right)$ is the predictive density, which is obtained by smoothing the forecast samples using a kernel density estimator.[15]

Some authors argue that the LPDS is sensitive to extreme observations and propose the continuous ranked probability score (CRPS) as a robust alternative (see, e.g., Ravazzolo and Vahey, 2014; Clark and Ravazzolo, 2015, among others). Hence I also compute the CRPS and document its average across evaluation periods:

---

[14]For MAFE as a loss function the predictive median would be the optimal forecast. But for comparison with other measures it seems preferable that the predictive mean is used throughout.

[15]I use the Matlab routine *ksdensity* with default options. The frequently used Gaussian approximation of Adolfson, Lindé, and Villani (2007) is not applicable since densities may be asymmetric.

$$\overline{\text{CRPS}}_{i,h}^{j} = \frac{1}{T_1 - T_0 + 1} \sum_{T=T_0+h}^{T_1+h} \text{CRPS}_{i,T|T-h}^{j}, \tag{9}$$

with

$$\text{CRPS}_{i,T|T-h}^{j} = \int_{-\infty}^{\infty} \left(P_{i,T|T-h}^{j}(y) - \mathbb{I}\left[y \geq y_{i,T} \right] \right)^2 dy, \tag{10}$$

where $P_{i,T|T-h}^{j}(\cdot)$ is the cumulative density function (CDF) to $p_{i,T|T-h}^{j}(\cdot)$. The CDF of the realization is given by the indicator function $\mathbb{I}\left[y \geq y_{i,T} \right]$ taking value 1 if $y \geq y_{i,T}$ and 0 otherwise.

In practice, I follow Gneiting and Raftery (2007) and calculate the CRPS using its alternative representation:

$$\text{CRPS}_{i,T|T-h}^{j} = \text{E}|Y - y_{i,T}| - \frac{1}{2}\text{E}|Y - Y'|, \tag{11}$$

where Y and Y' are independent draws from $p_{i,T|T-h}^{j}(\cdot)$. The expectation terms (E) are approximated using the forecast samples.[16]

Since the LPDS and CRPS are both silent about the absolute density forecast accuracy, I again report them for each specification relative to the benchmark:

$$\left(\overline{\text{LPDS}}_{i,h}^{j} - \overline{\text{LPDS}}_{i,h}^{B} \right) \quad \text{and} \quad \left(\overline{\text{CRPS}}_{i,h}^{j} / \overline{\text{CRPS}}_{i,h}^{B} \right). \tag{12}$$

Finally, I assess the absolute calibration of the predictive densities using the probability integral transform (PIT). For a given predictive density, the PIT is the corresponding CDF evaluated at the realization:

$$\text{PIT}_{i,T|T-h}^{j} = P_{i,T|T-h}^{j}(y_{i,T}). \tag{13}$$

If the proposed predictive density is consistent with the (unobserved) true density, the sequence $\left\{ \text{PIT}_{i,T|T-h}^{j} \right\}_{T=T_0+h}^{T_1+h}$ has an uniform distribution. I follow Diebold, Gunther, and Tay (1998) and provide a graphical assessment of the PITs. In particular, I divide the unit interval into $\hat{n} = 5$ equally sized bins and depict the number of PITs falling into any bin. If the predictive density would be calibrated correctly, each bin should contain $\hat{p} = 100\%/\hat{n} = 20\%$ of the PITs. To detect any significant departure from uniformity, I compute 90% confidence bands utilizing a normal approximation, i.e. $\hat{p} \pm 1.645\sqrt{\hat{p}(1-\hat{p})/(T_1 - T_0 + 1)}$ (see also Rossi and Sehkposyan, 2014).

---

[16]Note that the CRPS can be viewed as the distributional analogue of the MAFE, since for $Y = Y'$, i.e. for point forecasts, the CRPS is the same as the MAFE (see also Arora, Little, and McSharry, 2013).

For all other forecast accuracy measures I consider the test of unconditional predictive ability of Giacomini and White (2006) to check for significance at the 1%, 5%, and 10% level. Asymptotic standard errors are calculated with the Newey and West (1987) method that accounts for heteroskedasticity and autocorrelation in forecast errors. While I could provide evidence for all variables, I restrict the forecast comparison to four major macroeconomic series: GDP growth, the unemployment rate, CPI inflation, and the treasury rate.[17] [18]

4.3 Point Forecasts

I begin the forecast comparison by analyzing the MFE, which are reported for each specification and series in Table 4. In each case, the MFE is in percentage points (pp). The table shows that the benchmark and threshold specification overpredict GDP growth during the zero lower bound period, but not significantly. Replacing the federal funds rate by its shadow rate reinforces this bias, as I would expect additional stimulus to do. Simulating the BVAR with $\log(R_t)$ rather than R_t also inflates the MFE. At $h = 4$ the logarithmic specification significantly overpredicts GDP growth by 1.3 pp. The small movements in the federal funds rate are large once the logarithm is taken, leading to substantial increases in growth forecasts that are not supported by the data. Moreover, both policy shock specifications show MFE that are similar to those of the benchmark. Restricting the policy rate from below and/or above does not affect the location of the predictive mean in this case. In contrast, the soft condition specification produces point forecasts that are significantly above actual growth rates at all horizons. This bias could be the result of a large number of positive demand shocks added to the BVAR to prevent the policy rate from falling below 0%. Indeed, if the soft condition specification is combined with external information, and the size of the demand shocks limited, the MFE are close to zero at $h = 2$ and later.

With respect to the unemployment rate, all specifications show MFE that are small and in no case significantly different from zero. Furthermore, I obtain that all of them tend to underpredict CPI inflation at short horizons, while their predictive means are quite accurate at $h = 4$; except for the soft condition specification, which produces point forecasts that are well above actual rates (but not significantly). This finding confirms the idea that the soft condition specification puts a large weight on positive demand shocks to ensure a nonnegative policy rate. As for GDP growth, the potential bias can be removed if this variant is coupled with external information.

---

[17]In fact, I could not provide a complete forecast comparison for the federal fund rate since the shadow rate BVAR does not include the policy instrument.

[18]Forecasts for GDP growth and CPI inflation are obtained by computing the difference of the respective level forecasts.

14

Table 4: Mean Forecast Errors

Specification	Forecast Horizon h (in Quarters)			
	1	2	3	4
GDP Growth				
Benchmark	0.42	0.23	0.25	0.40
Threshold	0.28	0.33	0.17	0.14
Shadow Rate	0.47	0.30	0.37	0.60
Logarithmic	0.77**	0.74	0.93	1.27**
Policy Shock	0.42	0.19	0.21	0.36
Policy Shock (EI)	0.42	0.21	0.22	0.35
Soft Condition	1.07***	0.78**	0.78**	0.79**
Soft Condition (EI)	0.36	0.04	−0.04	0.11
Unemployment Rate				
Benchmark	0.00	0.04	0.09	0.13
Threshold	0.03	0.10	0.18	0.25
Shadow Rate	0.00	0.03	0.06	0.08
Logarithmic	0.01	0.03	0.01	−0.04
Policy Shock	0.00	0.04	0.08	0.12
Policy Shock (EI)	0.00	0.04	0.08	0.13
Soft Condition	−0.04	−0.08	−0.12	−0.15
Soft Condition (EI)	0.00	0.05	0.12	0.20
CPI Inflation				
Benchmark	−0.51	−0.62	−0.37	0.01
Threshold	−0.82	−0.39	0.04	0.03
Shadow Rate	−0.62	−0.78	−0.56	−0.15
Logarithmic	−0.53	−0.78	−0.56	−0.18
Policy Shock	−0.51	−0.63	−0.37	0.08
Policy Shock (EI)	−0.51	−0.64	−0.41	−0.04
Soft Condition	−0.19	−0.30	0.03	0.54
Soft Condition (EI)	−0.37	−0.56	−0.41	−0.08
Treasury Rate				
Benchmark	0.00	0.02	0.10	0.23
Threshold	−0.07	−0.12	−0.09	−0.07
Shadow Rate	0.00	0.02	0.10	0.22
Logarithmic	0.01	0.03	0.08	0.15
Policy Shock	0.05	0.09	0.17	0.29
Policy Shock (EI)	0.01	0.01	0.06	0.15
Soft Condition	0.20***	0.35***	0.48**	0.63**
Soft Condition (EI)	−0.01	−0.04	−0.02	0.07

Notes: this table shows the mean forecast errors for the period 2009:h to 2014:h. ***, **, and * indicate that the errors are significantly different from zero at the 1%, 5%, and 10% level, respectively.

Table 5: Root Mean Squared Forecast Errors

| Specification | Forecast Horizon h (in Quarters) | | | |
	1	2	3	4
	GDP Growth			
Benchmark	2.25	2.25	2.37	2.11
Threshold	1.12	1.32*	1.28**	1.32**
Shadow Rate	1.01	1.02*	1.03***	1.04**
Logarithmic	1.06	1.09	1.17	1.29*
Policy Shock	1.00	1.01	1.01	1.02
Policy Shock (EI)	1.00	1.02	1.02	1.01
Soft Condition	1.10	0.98	0.97	1.01
Soft Condition (EI)	1.03	0.91*	0.90	0.92*
	Unemployment Rate			
Benchmark	0.25	0.42	0.50	0.62
Threshold	0.99	1.19**	1.40***	1.51***
Shadow Rate	0.98**	0.98**	0.97*	0.96*
Logarithmic	0.99	1.06	1.17*	1.26*
Policy Shock	1.00	1.01	1.02	1.02
Policy Shock (EI)	1.00	1.01	1.02	1.02
Soft Condition	1.17	1.21	1.27	1.26
Soft Condition (EI)	1.20	1.25	1.35	1.38*
	CPI Inflation			
Benchmark	2.20	2.61	2.28	1.75
Threshold	1.62	1.31	1.25	1.31
Shadow Rate	1.01	1.00	0.98	0.96
Logarithmic	1.05	1.12	1.22	1.26
Policy Shock	1.00	1.02	1.03	0.99
Policy Shock (EI)	1.00	1.02	1.03	0.98
Soft Condition	0.96	0.93	0.89	0.93
Soft Condition (EI)	0.98	0.95	0.92	0.89
	Treasury Rate			
Benchmark	0.42	0.74	0.91	1.08
Threshold	0.88	0.81	0.74	0.73
Shadow Rate	1.02	1.02	1.02	1.03
Logarithmic	0.97	0.93	0.93	0.95
Policy Shock	0.96	0.92	0.94	0.97
Policy Shock (EI)	0.94	0.89*	0.90*	0.91**
Soft Condition	1.05	0.92	0.91	0.93
Soft Condition (EI)	0.96	0.80	0.74	0.72*

Notes: this table shows the root mean squared forecast errors for the period 2009:h to 2014:h. The errors are provided in levels for the benchmark and as ratios relative to the benchmark for the other specifications. A ratio below unity means that the respective specification outperforms the benchmark. ***, **, and * indicate that errors are significantly different from each other at the 1%, 5%, and 10% level, respectively.

Finally, almost all point forecasts are correctly calibrated for the treasury rate. The soft condition specification is again an exception, showing significantly positive MFE at all horizons. If combined with external information, however, this specification produces point forecasts that are close to actual values on average, suggesting that conditioning on a plausible path for the federal funds rate is profitable in this case. Similarly, the policy shock specification with external information is also quite accurate on average.

I continue the forecast comparison by discussing the RMSFE, which are reported in Table 5 for the benchmark in levels (pp) and the other specifications as ratios relative to the benchmark. A ratio below unity means that a specification outperforms the benchmark. The table suggests that replacing the federal funds rate by its shadow rate delivers RMSFE for GDP growth that are slightly above those of the benchmark. In contrast, the threshold and logarithmic approach show RMSFE ratios that are much larger and also increase with the forecast horizon. At $h = 4$ their RMSFE are about 30% higher as compared to the benchmark. Moreover, both policy shock specifications produce similar RMSFE ratios of around unity. And despite its biased predictive mean, the soft condition specification does not do significantly worse than the benchmark. In fact, if coupled with external information, the soft condition specification is the only model that delivers RMSFE ratios below unity at $h = 2$ and later.

For the unemployment rate I detect a small but significant improvement of the shadow rate specification over the benchmark at all horizons. In contrast, both soft condition specifications inflate the RMSFE a lot for the unemployment rate, while performing satisfactory for inflation. The threshold and logarithmic specification do significantly worse for the unemployment rate at longer horizons, but their RMSFE ratios are not significantly larger than unity for inflation. Furthermore, the policy shock specification, with and without external information, is not significantly different from the benchmark in terms of RMSFE for both variables.

With respect to the treasury rate, I obtain that almost all variants produce RMSFE ratios that are below unity at most horizons. Conditioning on financial market expectations is particularly helpful, leading to small but significant gains over the benchmark when implemented through policy shocks and larger ones when the soft condition approach is used. In this case, the RMSFE is cut by 28% at $h = 4$. Only the shadow rate specification increases the RMSFE in comparison to the benchmark, though not significantly.

The ratios based on MAFE are reported in Appendix B. By and large, the results are qualitatively similar, which suggests that the RMSFE are not distorted by large forecast errors.

4.4 Density Forecasts

Next I extend the forecast comparison to the predictive density by presenting the LPDS in Table 6. The scores are in levels for the benchmark and in differences to the benchmark for the other specifications. A positive difference means that a specification outperforms the benchmark. The table reveals that all specifications produce a lower LPDS for GDP growth than the benchmark, except the soft condition specification with external information which delivers higher scores. At $h = 4$ its LPDS is 7% higher and the difference significant. Thanks to an accurately calibrated predictive mean and a small variance, this variant attaches a relatively high probability to actual growth rates. In contrast, the logarithmic and threshold specification show scores that are about 20% lower than that of the benchmark at $h = 4$, which is consistent with their poor performance in terms of RMSFE, while the LPDS also penalizes their relatively large variance.

For the unemployment rate I obtain that both soft condition specifications deliver LPDS that are much smaller than those of the benchmark. At $h = 4$ their score is 33% lower with external information and 12% without, while the former difference is also significant. Moreover, the logarithmic specification is as accurate as the benchmark at $h = 1$, but its density forecast accuracy is deteriorating over time, showing an underperformance of 18% for the longest horizon. The logarithmic specification often shows this pattern of relatively accurate forecasts at $h = 1$ and poor performance at longer horizons, suggesting that the nonlinear relation between the federal funds rate and other macroeconomic variables is not supported by data, inflating forecast errors over time. Similarly, I again observe a poor performance of the threshold BVAR. Moreover, I find that both policy shock specifications produce scores that are a bit lower, while the shadow rate specification shows small insignificant gains. For CPI inflation the LPDS are similar across specifications. Only the threshold and logarithmic specification show significantly lower scores.

Combining the policy shock and soft condition specification with federal funds rate expectations is also profitable for the treasury rate in terms of LPDS. Both specifications produce larger scores than the benchmark, and this benefit increases with the forecast horizon. At $h = 4$ they significantly improve upon the benchmark by 8% and 22%, respectively. For all other variants I obtain scores that are similar to those of the benchmark.

Before I assess the density forecast accuracy of the different variants in absolute terms using PIT histograms, I briefly comment on the CRPS, which are reported in Appendix B. Overall, the ratios are similar to those obtained under the LPDS, suggesting that the scores are not distorted by outliers, but are a reasonable description of the density forecast accuracy of the specifications.

I conclude the forecast comparison by inspecting the PIT histograms for the longest horizon in Figure 3. This proceeding is useful since the shape of the histograms allows me to explore the

Table 6: Average Log Predictive Density Scores

Specification	1	2	3	4
		Forecast Horizon h (in Quarters)		
		GDP Growth		
Benchmark	−2.26	−2.25	−2.32	−2.25
Threshold	−0.18*	−0.29**	−0.29***	−0.28***
Shadow Rate	0.00	−0.02	−0.02*	−0.03***
Logarithmic	−0.05	−0.09*	−0.16**	−0.21***
Policy Shock	0.00	−0.01	0.00	−0.02***
Policy Shock (EI)	0.00	−0.01**	−0.01	−0.01
Soft Condition	−0.08	0.00	0.00	0.00
Soft Condition (EI)	0.03	0.07*	0.08	0.07**
		Unemployment Rate		
Benchmark	0.07	−0.42	−0.66	−0.92
Threshold	−0.05	−0.28*	−0.38***	−0.41***
Shadow Rate	0.02	0.02	0.03	0.03
Logarithmic	0.01	−0.08	−0.14	−0.18*
Policy Shock	0.00	−0.01**	−0.01	0.00
Policy Shock (EI)	0.00	0.00	−0.01	−0.01
Soft Condition	−0.16	−0.19	−0.13	−0.12
Soft Condition (EI)	−0.23	−0.29	−0.31*	−0.33*
		CPI Inflation		
Benchmark	−2.05	−2.14	−2.05	−2.00
Threshold	−0.25***	−0.18***	−0.24***	−0.29***
Shadow Rate	0.00	0.01	0.01	0.02
Logarithmic	−0.05	−0.09**	−0.12**	−0.10***
Policy Shock	0.00	0.01	0.01	0.00
Policy Shock (EI)	0.00	−0.01	0.00	0.01
Soft Condition	0.03	0.04	0.03	0.00
Soft Condition (EI)	0.02	0.03	0.02	0.05
		Treasury Rate		
Benchmark	−0.54	−1.06	−1.28	−1.45
Threshold	0.00	−0.02	0.02	−0.01
Shadow Rate	−0.01	0.00	0.01	−0.01
Logarithmic	0.01	0.00	−0.03	−0.05
Policy Shock	0.03	0.04	0.02	0.00
Policy Shock (EI)	0.05	0.07*	0.08***	0.08***
Soft Condition	−0.04	−0.01	−0.02	−0.03
Soft Condition (EI)	0.05	0.12	0.18	0.22**

Notes: this table shows the average log predictive density scores for the period 2009:h to 2014:h. The scores are provided in levels for the benchmark and in differences to the benchmark for the other specifications. A positive difference means that the respective specification outperforms the benchmark. ***, **, and * indicate that scores are significantly different from each other at the 1%, 5%, and 10% level.

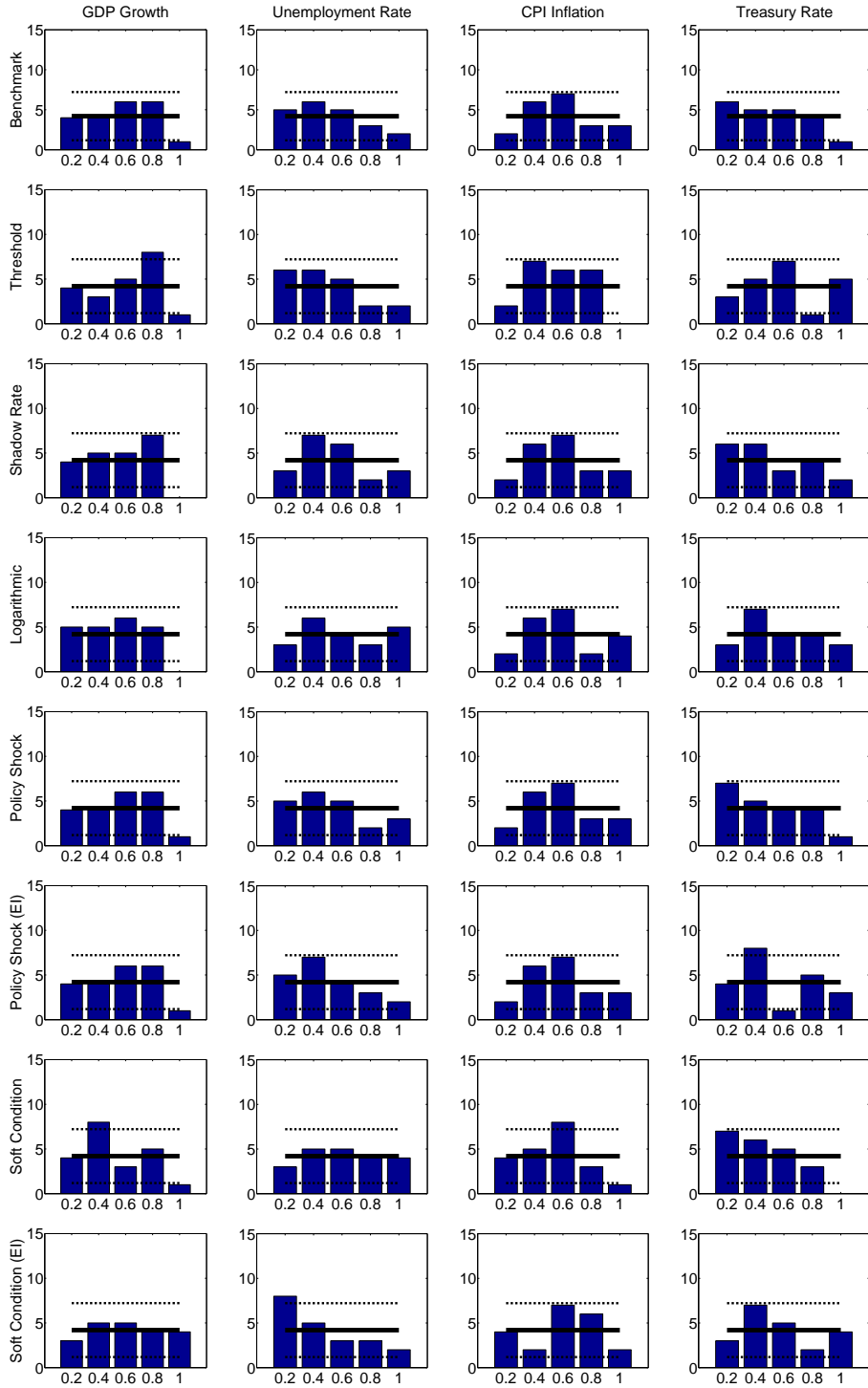Figure 3: Probability Integral Transform Histograms ($h = 4$). Notes: this figure shows the probability integral transform histograms for $h = 4$ and the period 2009:4 to 2014:4. The solid line denotes uniformity. The dashed lines are a 90% confidence band. x-axis: quantile; y-axis: number of realizations.

reasons behind the poor density forecast accuracy of some specifications at $h = 4$. The plots for $h = 1$ are provided in Appendix B and are not further discussed here because differences across specifications are small in this case.

Figure 3 reveals that all variants significantly overpredict GDP growth at $h = 4$, except the soft condition specification with external information which shows correct calibration. The realistic calibration of this specification is in line with its low MFE and relatively strong performance in terms of RMSFE and LPDS. However, I obtain the opposite for the unemployment rate. While I cannot reject correct calibration for any other variant, this soft condition specification attaches a too large probability to high unemployment rates, which could be an explanation for its poor LPDS. The PITs also suggest that most specifications are correctly calibrated for CPI inflation, but the threshold and soft condition specification are not. Finally, the PITs detect a poor calibration of the benchmark with respect to the treasury rate. At $h = 4$ the benchmark significantly overpredicts the series, while replacing the federal funds rate by its shadow rate or logarithm helps to achieve correct calibration. Also, I cannot reject correct calibration for the soft condition specification if coupled with external information.

5   Summary and Conclusion

In this paper I discuss how the forecast accuracy of a BVAR is affected by introducing the zero lower bound on the federal funds rate. As a benchmark I adopt a common BVAR specification, including 18 variables in (log-)levels, estimated shrinkage, and no nonlinearity. Then I entertain alternative specifications of the zero lower bound. I account for the possibility that the effect of monetary policy on the economy is different in a low policy rate regime, replace the federal funds rate by its shadow rate, consider a logarithmic transformation, feed in monetary policy shocks, or utilize conditional forecasts allowing for all shocks implemented through a rejection sampler. The latter two are also coupled with interest rate expectations from future contracts. The comparison is based on the accuracy of point and density forecasts of GDP growth, the unemployment rate, CPI inflation, and the treasury rate during the period 2009:1 to 2014:4. For a robust assessment I consider several relative and absolute forecast metrics.

I obtain a large set of empirical results in this paper. To condense the material, I provide a qualitative summary of them in Tables 7 and 8 for the shortest and longest forecast horizon, respectively. My interpretation of the results is as follows. First, the introduction of the zero lower bound hardly affects the forecast accuracy of the BVAR with respect to real activity and prices at $h = 1$ but does at $h = 4$. While this is by construction for the policy shock specifications, it is also observed for the other specifications. A practical forecaster should thus be more

concerned about the zero lower bound when she is interested in long-term forecasts. Second, the performance of the alternative specifications differs at $h = 4$, suggesting that this modeling choice is not innocuous. In particular, the introduction of the zero lower bound is not beneficial *per se*, but it depends on how it is done and which series is forecasted.

With caution, I recommend the shadow rate specification. Except for GDP growth, this specification delivers a satisfactory forecast ability in absolute terms and relative to the benchmark. In contrast, the threshold and logarithmic specifications are no good choices, while both policy shock specifications are as accurate as the benchmark. Finally, the soft condition specification is useful only in combination with external information, but then helps to achieve a better calibration of predictive densities and reduces forecast errors relative to the benchmark for all series, except the unemployment rate. Since the policy rate will remain low for some time, these findings could prove useful for practical forecasters.

Table 7: Summary of Results ($h = 1$)

Specification	MFE	RMSFE	MAFE	LPDS	CRPS	PITS
			Forecast Metric			
GDP Growth						
Benchmark	✓					✓
Threshold	✓			✗	✗	✓
Shadow Rate	✓					✓
Logarithmic	✗					✓
Policy Shock	✓					✓
Policy Shock (EI)	✓					✓
Soft Condition	✗					✗
Soft Condition (EI)	✓					✓
Unemployment Rate						
Benchmark	✓					✗
Threshold	✓					✓
Shadow Rate	✓	✓	✓		✓	✓
Logarithmic	✓					✓
Policy Shock	✓					✗
Policy Shock (EI)	✓					✗
Soft Condition	✓					✓
Soft Condition (EI)	✓					✓
CPI Inflation						
Benchmark	✓					✓
Threshold	✓			✗		✓
Shadow Rate	✓					✓
Logarithmic	✓					✗
Policy Shock	✓					✓
Policy Shock (EI)	✓					✓
Soft Condition	✓					✓
Soft Condition (EI)	✓					✓
Treasury Rate						
Benchmark	✓					✗
Threshold	✓					✓
Shadow Rate	✓					✓
Logarithmic	✓					✓
Policy Shock	✓					✗
Policy Shock (EI)	✓				✓	✓
Soft Condition	✗					✗
Soft Condition (EI)	✓					✓

Notes: this table shows a summary of the results for $h = 1$. Absolute metrics: ✓ = correct calibration, ✗ = incorrect calibration. Relative metrics: ✓ = gains over benchmark; ✗ = losses over benchmark; blank = no gains/losses over benchmark. Significance level is 10% in each case.

Table 8: Summary of Results ($h = 4$)

Specification	Forecast Metric					
	MFE	RMSFE	MAFE	LPDS	CRPS	PITS
GDP Growth						
Benchmark	✓					✗
Threshold	✓	✗	✗	✗	✗	✗
Shadow Rate	✓	✗		✗	✗	✗
Logarithmic	✗	✗	✗	✗	✗	✗
Policy Shock	✓			✗	✗	✗
Policy Shock (EI)	✓					✗
Soft Condition	✗					✗
Soft Condition (EI)	✓	✓		✓		✓
Unemployment Rate						
Benchmark	✓					✓
Threshold	✓	✗	✗	✗	✗	✓
Shadow Rate	✓	✓			✓	✓
Logarithmic	✓	✗	✗	✗		✓
Policy Shock	✓					✓
Policy Shock (EI)	✓					✓
Soft Condition	✓		✗			✓
Soft Condition (EI)	✓	✗		✗	✗	✗
CPI Inflation						
Benchmark	✓					✓
Threshold	✓			✗	✗	✗
Shadow Rate	✓		✓			✓
Logarithmic	✓			✗		✓
Policy Shock	✓					✓
Policy Shock (EI)	✓		✓			✓
Soft Condition	✓					✗
Soft Condition (EI)	✓					✓
Treasury Rate						
Benchmark	✓					✗
Threshold	✓					✗
Shadow Rate	✓		✗			✓
Logarithmic	✓					✓
Policy Shock	✓					✗
Policy Shock (EI)	✓	✓		✓	✓	✗
Soft Condition	✗					✗
Soft Condition (EI)	✓	✓		✓	✓	✓

Notes: this table shows a summary of the results for $h = 4$. Absolute metrics: ✓ = correct calibration, ✗ = incorrect calibration. Relative metrics: ✓ = gains over benchmark; ✗ = losses over benchmark; blank = no gains/losses over benchmark. Significance level is 10% in each case.

References

Adolfson, M., J. Lindé, and M. Villani (2007). Forecasting Performance of an Open Economy DSGE Model. *Econometric Reviews 26*(2-4), 289–328.

Arora, S., M. A. Little, and P. E. McSharry (2013). Nonlinear and Nonparametric Modeling Approaches for Probabilistic Forecasting of the US Gross National Product. *Studies in Nonlinear Dynamics & Econometrics 17*(4), 395–420.

Bańbura, M., D. Giannone, and M. Lenza (2015). Conditional Forecasts and Scenario Analysis with Vector Autoregressions for Large Cross-Sections. *International Journal of Forecasting 31*(3), 739–756.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian Vector Autoregressions. *Journal of Applied Econometrics 25*(1), 71–92.

Berg, T. O. and S. R. Henzel (2015). Point and Density Forecasts for the Euro Area Using Bayesian VARs. *International Journal of Forecasting 31*(4), 1067–1095.

Bernanke, B. S. (2013). Communication and Monetary Policy. Speech at the National Economists Club Annual Dinner, Herbert Stein Memorial Lecture, Washington, D.C., November 19.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Bayesian VARs: Specification Choices and Forecast Accuracy. *Journal of Applied Econometrics 30*(1), 46–73.

Clark, T. E. and M. W. McCracken (2014). Evaluating Conditional Forecasts from Vector Autoregressions. Working Paper 1413, Federal Reserve Bank of Cleveland.

Clark, T. E. and F. Ravazzolo (2015). Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Cogley, T. and T. J. Sargent (2002). Evolving Post-World War II U.S. Inflation Dynamics. In B. S. Bernanke and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2001*, Volume 16, pp. 331–388. Cambridge: MIT Press.

Del Negro, M. and F. Schorfheide (2013). DSGE Model-Based Forecasting. In G. Elliot and A. Timmerman (Eds.), *Handbook of Economic Forecasting*, Volume 24, pp. 57–140. North-Holland: Elsevier.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review 39*(4), 863–883.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews 3*(1), 1–100.

Gelman, A., J. Hwang, and A. Vehtari (2014). Understanding Predictive Information Criteria for Bayesian Models. *Statistics and Computing 24*(6), 997–1016.

Geweke, J. and G. Amisano (2010). Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. *International Journal of Forecasting 26*(2), 216–230.

Giacomini, R. and H. White (2006). Tests of Conditional Predicitve Ability. *Econometrica 74*(6), 1545–1578.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior Selection for Vector Autoregressions. *Review of Economics and Statistics 97*(2), 412–435.

Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Hubrich, K. and T. Teräsvirta (2013). Thresholds and Smooth Transitions in Vector Autoregressive Models. In T. B. Fomby, L. Kilian, and A. Murphy (Eds.), *Advances in Econometrics*, Volume 32, pp. 273–326. Bingley: Emerald Group Publishing Limited.

Kadiyala, K. R. and S. Karlsson (1993). Forecasting with Generalized Bayesian Vector Autoregressions. *Journal of Forecasting 12*(4), 365–378.

Kadiyala, K. R. and S. Karlsson (1997). Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics 12*(2), 99–132.

Koop, G. (2013). Forecasting with Medium and Large Bayesian VARS. *Journal of Applied Econometrics 28*(2), 177–203.

Krippner, L. (2013). Measuring the Stance of Monetary Policy in Zero Lower Bound Environments. *Economics Letters 118*(1), 135–138.

Litterman, R. B. (1980). A Bayesian Procedure for Forecasting with Vector Autoregressions. Manuscript, Massachusetts Institute of Technology, Department of Economics.

Litterman, R. B. (1986). Forecasting With Bayesian Vector Autoregressions - Five Years of Experience. *Journal of Business and Economic Statistics 4*(1), 25–38.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Ravazzolo, F. and S. P. Vahey (2014). Forecast Densities for Economic Aggregates from Disaggregate Ensembles. *Studies in Nonlinear Dynamics & Econometrics 18*(4), 367–381.

Rogoff, K. S. (2015). Costs and Benefits to Phasing out Paper Currency. In J. A. Parker and M. Woodford (Eds.), *NBER Macroeconomics Annual 2014*, Volume 29, pp. 445–456. Chicago: University of Chicago Press.

Rossi, B. and T. Sehkposyan (2014). Evaluating Predictive Densities of US Output Growth and Inflation in a Large Macroeconomic Data Set. *International Journal of Forecasting 30*(3), 662–682.

Sims, C. A. (1993). A Nine-Variable Probabilistic Macroeconomic Forecasting Model. In J. H. Stock and M. W. Watson (Eds.), *Business Cycles, Indicators and Forecasting*, pp. 179–212. Chicago: University of Chicago Press.

Smets, F., A. Warne, and R. Wouters (2014). Professional Forecasters and Real-Time Forecasting with a DSGE Model. *International Journal of Forecasting 30*(4), 981–995.

Waggoner, D. F. and T. Zha (1999). Conditional Forecasts in Dynamic Multivariate Models. *Review of Economics and Statistics 81*(4), 639–651.

Weiss, A. A. (1996). Estimating Time Series Models Using the Relevant Cost Function. *Journal of Applied Econometrics 11*(5), 539–560.

Wu, J. C. and F. D. Xia (2015). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. forthcoming in *Journal of Money, Credit, and Banking*.

A Implementation and Estimation Details

In this appendix I provide the implementation and estimation details for the BVAR in Section 2. Moreover, I explain how the algorithm needs to be modified for some alternative specifications of the zero lower bound and how forecasts are obtained.

A.1 Implementing the Priors

Consider the VAR from Section 2:

$$y_t = c + B_1 y_{t-1} + ... + B_p y_{t-p} + u_t, \quad u_t \sim \mathrm{N}\left(0, \Sigma\right), \tag{A.1}$$

which can be written compactly as

$$y = xB + u, \tag{A.2}$$

where $y = (y_1, ..., y_T)'$, $x_t = \left(y'_{t-1}, ..., y'_{t-p}, 1\right)'$, $x = (x_1, ..., x_T)'$, $B = (B_1, ..., B_p, c)'$, and $u = (u_1, ..., u_T)'$. Finally, let $\beta = \mathrm{vec}\left(B\right)$ with $\mathrm{vec}\left(\cdot\right)$ being the column stacking operator.

First, I discuss the implementation of the Minnesota prior on (β, Σ). In particular, I consider a conjugate Normal-Inverse-Wishart prior of the following form:

$$\Sigma \sim \mathrm{IW}\left(\Psi, d\right) \quad \text{and} \quad \beta | \Sigma \sim \mathrm{N}\left(b, \Sigma \otimes \Omega\right), \tag{A.3}$$

where $\otimes$ denotes the Kronecker product and the elements Ψ, d, b, and Ω are functions of hyperparameters. The conjugate prior implies a likelihood and posterior that come from the same family of distributions and hence makes Bayesian inference feasible even for large n.

Bańbura et al. (2010) implement the prior by constructing a set of dummy observations:

$$y^+ = \begin{bmatrix} \mathrm{diag}\left(\delta\sigma_1, ..., \delta\sigma_n\right)/\lambda_1 \\ 0_{n(p-1)\times n} \\ \mathrm{diag}\left(\sigma_1, ..., \sigma_n\right) \\ 0_{1\times n} \end{bmatrix}, \quad x^+ = \begin{bmatrix} \mathrm{diag}\left(1, 2, .., p\right) \otimes \mathrm{diag}\left(\sigma_1, ..., \sigma_n\right)/\lambda_1 & 0_{np\times 1} \\ 0_{n\times np} & 0_{n\times 1} \\ 0_{1\times np} & \epsilon \end{bmatrix}, \quad \text{(A.4)}$$

where $\mathrm{diag}\left(\cdot\right)$ denotes a diagonal matrix. The σ_j's account for the different scale and variability of the series and are set to the standard deviation of a residual from a univariate autoregression for the variable $y_{j,t}$ using the same lag length as in the VAR. The parameter ϵ is set to a small number (10^{-4}), reflecting a diffuse prior for the intercepts. Finally, the parameter λ_1 governs the

degree of shrinkage and hence the tightness of the prior. As $\lambda_1 \to \infty$ the prior becomes uninformative and posterior expectations coincide with the ordinary least squares (OLS) estimates. For $\lambda_1 \to 0$ the posterior approaches the dogmatic prior. λ_1 is hence a key parameter in the BVAR and its selection is discussed in detail below.

Second, the "sum-of-coefficients" prior is implemented by augmenting the system with n dummy observations and reflects the belief that a no-change forecast is a good forecast at the beginning of a sample period. In particular, I construct:

$$y^{++} = \mathrm{diag}\left(\delta\mu_1, ..., \delta\mu_n\right)/\lambda_2 , \qquad x^{++} = \left[1_{1\times p} \otimes \mathrm{diag}\left(\delta\mu_1, ..., \delta\mu_n\right) \quad 0_{n\times 1}\right]/\lambda_2 , \qquad \text{(A.5)}$$

where the μ_j's are set equal to the average value of $y_{j,t}$ over the estimation sample. Finally, the parameter λ_2 reflects the tightness of the prior. As $\lambda_2 \to 0$ the model can be expressed in terms of differenced data, there are as many unit roots as variables and there is no cointegration.

Third, the "dummy-initial-observation" prior introduces a single dummy observation and expresses the belief that a no-change forecast for all variables is a good forecast at the beginning of a sample period. The additional observation is

$$y^{+++} = \left(\delta\mu_1, ..., \delta\mu_n\right)/\lambda_3 , \qquad x^{+++} = \left[1_{1\times p} \otimes \left(\delta\mu_1, ..., \delta\mu_n\right) \quad 1\right]/\lambda_3 . \qquad \text{(A.6)}$$

The parameter λ_3 controls the tightness of the prior. As $\lambda_3 \to 0$ all variables are either stationary and at their unconditional mean, or nonstationary without drift and cointegrated.

The artificial observations are added on top of the actual data matrices, and the augmented regression model is used for inference:

$$y^* = x^* B + u^*, \qquad \text{(A.7)}$$

where $y^* = \left(y', y^{+'}, y^{++'}, y^{+++'}\right)'$, $x^* = \left(x', x^{+'}, x^{++'}, x^{+++'}\right)'$, $u^* = \left(u', u^{+'}, u^{++'}, u^{+++'}\right)'$, $u^{+'} = \left(u_1^+, ..., u_T^+\right)$, $u^{++'} = \left(u_1^{++}, ..., u_T^{++}\right)$, and $u^{+++'} = \left(u_1^{+++}, ..., u_T^{+++}\right)$. The latter three expressions are the corresponding residuals to y^+, y^{++}, and y^{+++}, respectively.

Let $\hat{\Sigma}$ and $\hat{\beta}$ be the covariance matrix and coefficients from an OLS regression of y^* on x^*. The conditional posterior distributions of the covariance matrix and coefficients can be computed in closed form as a function of the three shrinkage parameters which are collected in λ:

$$\Sigma|\lambda, y \sim \mathrm{IW}\left(\hat{\Sigma}, T + n + 2\right) \quad \text{and} \quad \beta|\Sigma, \lambda, y \sim \mathrm{N}\left(\hat{\beta}, \Sigma \otimes \left(x^{*'} x^*\right)^{-1}\right). \qquad \text{(A.8)}$$

In practice I get draws for the coefficients from the equivalent non-vectorized representation of their conditional posterior distribution:

$$B = \hat{B} + \text{chol}\left(\left(x^{*\prime}x^*\right)^{-1}\right) \times V \times \text{chol}\left(\Sigma\right)', \tag{A.9}$$

where $\text{chol}\left(\cdot\right)$ denotes the Cholesky decomposition and V is a $k \times n$ standard Gaussian matrix. Drawing coefficients from (A.9) rather than (A.8) is much faster for large n since the computation of the Kronecker product can be avoided (see also Carriero et al., 2015).

A.2   Choice for Shrinkage Parameters and MCMC Estimation

Following Giannone et al. (2015) I treat the shrinkage parameters as additional unknown parameters that can be estimated in an hierarchical fashion. The approach hence accounts for the uncertainty related to this specification choice and requires to add one more layer to the prior structure by placing a prior on the shrinkage parameters - a hyperprior $p\left(\lambda\right)$.

The marginal posterior for λ after integrating out the posterior uncertainty about the model's parameters is

$$p\left(\lambda|y\right) \propto p\left(y|\lambda\right) \cdot p\left(\lambda\right), \tag{A.10}$$

where $\propto$ denotes proportionality. The marginal likelihood of the model (conditional on λ) has a well-known analytic form (see also Carriero et al., 2015):

$$p\left(y|\lambda\right) = \zeta^{-1} \times |\Psi + \left(y - xb\right)'\left(I + x\Omega x'\right)^{-1}\left(y - xb\right)|^{-\frac{T+d}{2}}, \tag{A.11}$$

where

$$\zeta = \pi^{\frac{Tn}{2}} \times |\left(I + x\Omega x'\right)^{-1}|^{-\frac{n}{2}} \times |\Psi|^{-\frac{d}{2}} \times \frac{\Gamma_n\left(\frac{d}{2}\right)}{\Gamma_n\left(\frac{T+d}{2}\right)}, \tag{A.12}$$

and $\Gamma_n\left(\cdot\right)$ denoting the n-variate gamma function. The prior moments b, Ω and Ψ can be computed by OLS from the artificial data matrices, while $d = n + 2$.

The hyperprior needs to be chosen and reflects how confident we are about the values for λ. I follow common practice and choose a diffuse hyperprior. In particular, I consider a Gamma density with mode equal to 0.2 and standard deviation of 0.4 for λ_1, while for λ_2 and λ_3 the mode and standard deviation are both set equal to 1.

Since the joint posterior distribution for the parameters and λ is not available in closed form, Giannone et al. (2015) recommend a Metropolis-Hastings algorithm to simulate the distribution.

The sampler is a Markov chain Monte Carlo (MCMC) method and generates λ from its marginal posterior with a Metropolis update. After convergence of the sampler, the covariance matrix and coefficients can be drawn from their posterior distribution conditional on λ and forecasts computed in an iterative fashion. In detail, the algorithm works as follows.

Let N_B be the number of burn-in-draws and N_R the number of retained draws. The maximal forecast horizon is denoted by H. The steps are:

Step 1: Choose a starting point λ^0. I use the posterior mode, which is obtained by numerical optimization. For $j = 1, ..., N_B + N_R$ run a loop over the following steps.

Step 2: Draw a proposal λ^* from a jumping distribution $J\left(\lambda^*|\lambda^{j-1}\right) = N\left(\lambda^{j-1}, c \cdot \Sigma_m^2\right)$, where Σ_m^2 is the inverse of the Hessian computed at the posterior mode, and c is a scaling constant chosen to obtain an acceptance ratio of about 20%.

Step 3: Compute the acceptance ratio:

$$r = \frac{p\left(\lambda^*|y\right)}{p\left(\lambda^{j-1}|y\right)}, \tag{A.13}$$

where $p\left(\lambda|y\right)$ is given by Equation (A.10).

Step 4: Randomly draw ν from $U\left(0, 1\right)$.

Step 5: Accept or discard the proposal λ^* according to the following rule, and update, if necessary, the jumping distribution:

$$\lambda^j = \begin{cases} \lambda^* & : \text{ if } \nu \leq r \\ \lambda^{j-1} & : \text{ otherwise} \end{cases} \tag{A.14}$$

If $j \leq N_B$ repeat the previous four steps, otherwise continue.

Step 6: Draw $\Sigma^j|\lambda^j, y$ and $\beta^j|\Sigma^j, \lambda^j, y$ from their conditional posterior in (A.8) and (A.9).

Step 7: Generate $u_{T+1}^j, ..., u_{T+H}^j$ from $u_t \sim N\left(0, \Sigma^j\right)$.

Step 8: Compute the 1-step-ahead forecast as

$$\hat{y}_{T+1}^j = c^j + B_1^j y_T + ... + B_p^j y_{T-p+1} + u_{T+1}^j, \tag{A.15}$$

while the h-step-ahead forecasts are obtained by iteration:

$$\hat{y}_{T+h}^j = c^j + B_1^j \hat{y}_{T+h-1}^j + ... + B_p^j \hat{y}_{T+h-p}^j + u_{T+h}^j, \tag{A.16}$$

where $h = 1, ..., H$ and $\hat{y}_{T+h}^j = y_{T+h-p}$ for $h \leq p$. $\quad\square$

For some specifications the algorithm above needs to be modified in order to obtain a forecast sample that is consistent with a nonnegative federal funds rate. In particular, for the policy shock and soft condition specifications the final step is replaced by the following expressions:

Step 8 (*Policy Shock*): Compute the 1-step-ahead forecast as

$$\hat{y}^j_{T+1} = c^j + B^j_1 y_T + ... + B^j_p y_{T-p+1} + u^j_{T+1}. \tag{A.17}$$

If $\hat{R}^j_{T+1} \geq 0$ continue. Otherwise add a monetary policy shock (ϵ^j_{T+1}) such that the federal funds rate is equal to zero, i.e. $\epsilon^j_{T+1} = \left(-\hat{R}^j_{T+1}/\sigma^j_{n_1+1}\right) S^j e$, where $\sigma^j_{n_1+1}$ is the standard deviation of the shock, $S^j = \text{chol}\left(\Sigma^j\right)$, and $e = (0_{n_1} \ 1 \ 0_{n_2})'$. Obtain the h-step-ahead forecasts by iteration and always add a monetary policy shock (ϵ^j_{T+h}) if needed:

$$\hat{y}^j_{T+h} = c^j + B^j_1 \hat{y}^j_{T+h-1} + ... + B^j_p \hat{y}^j_{T+h-p} + u^j_{T+h}, \tag{A.18}$$

where $h = 1, ..., H$ and $\hat{y}^j_{T+h} = y_{T+h-p}$ for $h \leq p$. $\quad\square$

Step 8: (*Soft Condition*) Compute the 1-step-ahead forecast as

$$\hat{y}^j_{T+1} = c^j + B^j_1 y_T + ... + B^j_p y_{T-p+1} + u^j_{T+1}, \tag{A.19}$$

while the h-step-ahead forecasts are obtained by iteration:

$$\hat{y}^j_{T+h} = c^j + B^j_1 \hat{y}^j_{T+h-1} + ... + B^j_p \hat{y}^j_{T+h-p} + u^j_{T+h}, \tag{A.20}$$

where $h = 1, ..., H$ and $\hat{y}^j_{T+h} = y_{T+h-p}$ for $h \leq p$. If $\hat{R}^j_{T+h} \geq 0$ for $h = 1, ..., H$ accept the draw, otherwise repeat Steps 6 and 7 and use the alternative draw $\tilde{\Sigma}^j, \tilde{\beta}^j, \tilde{u}^j$ to compute forecasts. $\quad\square$

This algorithm yields $\left(\hat{y}^j_{T+1}, ..., \hat{y}^j_{T+H}\right)^{N_B+N_R}_{j=N_B+1}$ as a sample of forecasts generated from the joint posterior distribution for the parameters and λ, where $N_B = 5,000$ and $N_R = 25,000$.

Figure A.1 suggests that the posterior distribution for λ is well behaved and N_B sufficiently large to let the Markov chain converge to its ergodic random distribution. The figure also shows that the autocorrelation of the chain dies out quickly. Therefore I do not thin the chain to reduce dependence, but keep all draws for inference.
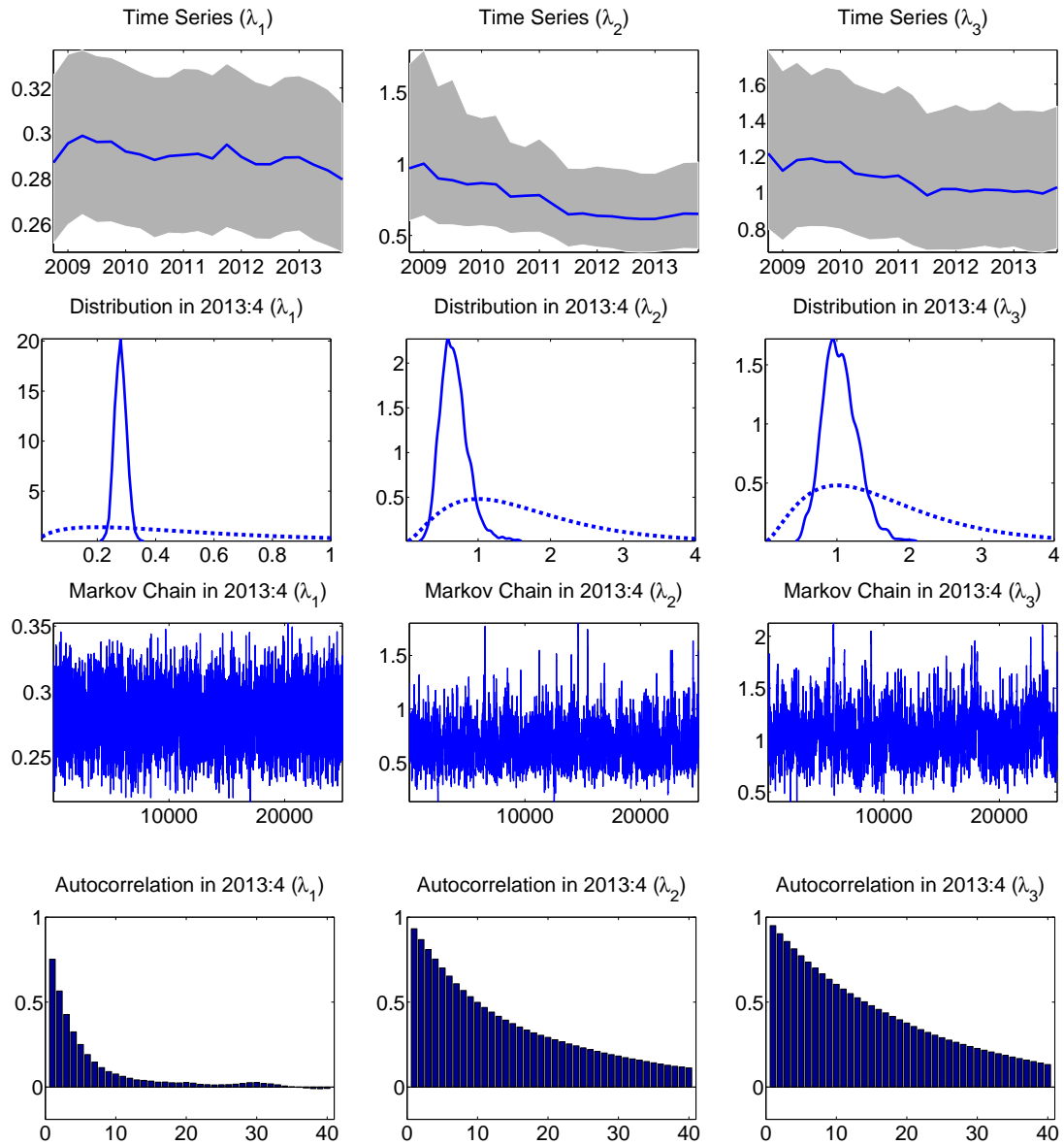
Figure A.1: Posterior Distribution of the Shrinkage Parameters. Notes: this figure shows the posterior distribution of the shrinkage parameters. First row: Median and 90% error band for all periods. Second row: Posterior distribution (solid) in final period and prior distribution (dashed). Third row: Markov chain in final period. Last row: Autocorrelation function in final period.

B Additional Material

In this appendix I provide some additional material to the paper. First, I present time series plots for the Wu-Xia shadow rate and the federal funds rate expectations (Figure B.1). Second, I report the MAFE and CRPS (Tables B.1 and B.2) as well as the PIT histograms at $h = 1$ (Figure B.2) for the BVAR in (log-)levels. Finally, I show the results for the BVAR in differences (Tables B.3 to B.5 and Figure B.3). From the latter we learn that the levels specification is indeed the appropriate benchmark since forecast errors are much larger on average for the difference specification. But the relative ranking of the BVAR specifications is hardly affected by this modeling choice.
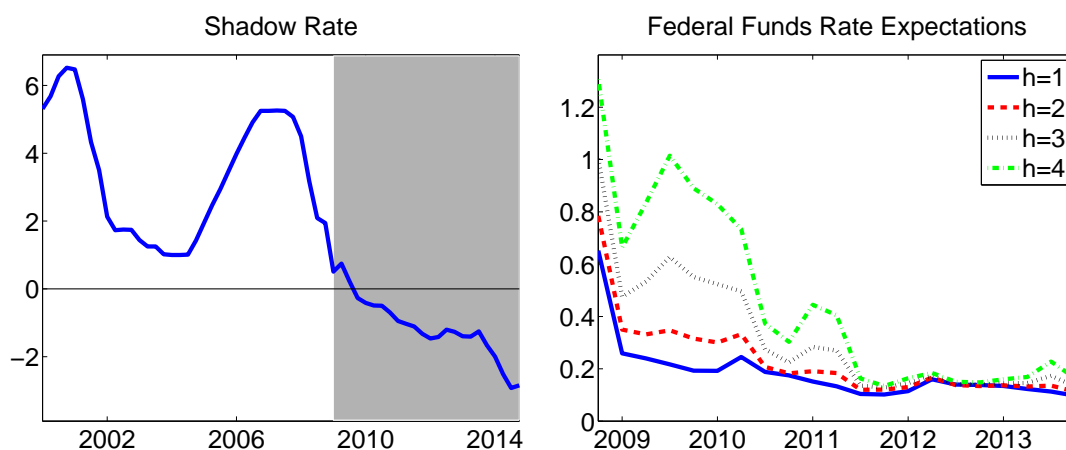

Figure B.1: Data. Notes: this figure shows the Wu-Xia shadow rate and the federal funds rate expectations from future contracts. The shaded area is the evaluation period. x-axis: year; y-axis: percent.

Table B.1: Mean Absolute Forecast Errors

Specification	Forecast Horizon h (in Quarters)			
	1	2	3	4
GDP Growth				
Benchmark	1.69	1.77	1.73	1.64
Threshold	1.21	1.38**	1.45***	1.35***
Shadow Rate	1.00	1.02	1.05**	1.04
Logarithmic	1.04	1.08	1.20	1.29*
Policy Shock	1.00	1.02	1.03	1.03
Policy Shock (EI)	1.00	1.02	1.03	1.03
Soft Condition	1.09	0.96	1.02	1.04
Soft Condition (EI)	1.00	0.91	0.91	0.95
Unemployment Rate				
Benchmark	0.18	0.30	0.38	0.49
Threshold	1.06	1.24**	1.42***	1.53***
Shadow Rate	0.98*	0.97**	0.95*	0.96
Logarithmic	1.06	1.10	1.25*	1.29*
Policy Shock	1.00	1.01**	1.01	1.01
Policy Shock (EI)	1.00	1.00	1.01	1.02
Soft Condition	1.04	0.98	1.05	1.07
Soft Condition (EI)	1.11	1.13	1.24*	1.30*
CPI Inflation				
Benchmark	1.54	1.74	1.51	1.39
Threshold	1.41	1.11	1.22	1.25
Shadow Rate	1.02	1.02	0.98	0.94***
Logarithmic	1.03	1.03	1.14	1.16
Policy Shock	1.00	1.01	1.02*	1.00
Policy Shock (EI)	1.00	1.01	1.03*	0.97*
Soft Condition	1.02	0.94	0.98	0.90
Soft Condition (EI)	1.03	0.98	1.01	0.88
Treasury Rate				
Benchmark	0.31	0.59	0.74	0.87
Threshold	0.93	0.85	0.70	0.71
Shadow Rate	1.01	1.00	1.01	1.03*
Logarithmic	1.00	0.90	0.91	0.95
Policy Shock	1.01	0.92	0.94	0.97
Policy Shock (EI)	0.96	0.91	0.90	0.92
Soft Condition	1.18	0.92	0.87	0.94
Soft Condition (EI)	1.00	0.79	0.75	0.71

Notes: this table shows the mean absolute forecast errors for the period 2009:h to 2014:h. The errors are provided in levels for the benchmark and as ratios relative to the benchmark for the other specifications. A ratio below unity means that the respective specification outperforms the benchmark. ***, **, and * indicate that errors are significantly different from each other at the 1%, 5%, and 10% level, respectively.

Table B.2: Average Continuous Ranked Probability Scores

Specification	Forecast Horizon h (in Quarters)			
	1	2	3	4
GDP Growth				
Benchmark	1.25	1.27	1.33	1.23
Threshold	1.18*	1.35**	1.37***	1.35***
Shadow Rate	1.01	1.01	1.03**	1.04**
Logarithmic	1.05	1.09	1.16*	1.26*
Policy Shock	1.00	1.01	1.01	1.02*
Policy Shock (EI)	1.00	1.01*	1.02	1.01
Soft Condition	1.08	0.98	0.99	1.00
Soft Condition (EI)	1.01	0.93	0.93	0.94
Unemployment Rate				
Benchmark	0.13	0.22	0.27	0.35
Threshold	1.02	1.23**	1.43***	1.51***
Shadow Rate	0.98**	0.97**	0.97*	0.96*
Logarithmic	1.03	1.10	1.19	1.24
Policy Shock	1.00	1.01	1.01	1.02
Policy Shock (EI)	1.00	1.00	1.01	1.02
Soft Condition	1.09	1.08	1.13	1.13
Soft Condition (EI)	1.14	1.19	1.30*	1.36**
CPI Inflation				
Benchmark	1.12	1.26	1.13	0.97
Threshold	1.44	1.22	1.23**	1.31**
Shadow Rate	1.01	1.00	0.98	0.97
Logarithmic	1.04	1.09	1.16	1.16
Policy Shock	1.00	1.01	1.02	1.00
Policy Shock (EI)	1.00	1.01	1.02	0.98
Soft Condition	0.98	0.96	0.94	0.98
Soft Condition (EI)	1.00	0.97	0.97	0.93
Treasury Rate				
Benchmark	0.24	0.41	0.50	0.59
Threshold	0.92	0.90	0.84	0.85
Shadow Rate	1.01	1.01	1.01	1.02
Logarithmic	0.98	0.95	0.97	0.99
Policy Shock	0.97	0.94	0.96	0.98
Policy Shock (EI)	0.95*	0.91*	0.91**	0.91**
Soft Condition	1.05	0.94	0.95	0.98
Soft Condition (EI)	0.95	0.83	0.79	0.76*

Notes: this table shows the average continuous ranked probability scores for the period 2009:h to 2014:h. The scores are provided in levels for the benchmark and as ratios relative to the benchmark for the other specifications. A ratio below unity means that the respective specification outperforms the benchmark. ***, **, and * indicate that scores are significantly different from each other at the 1%, 5%, and 10% level.
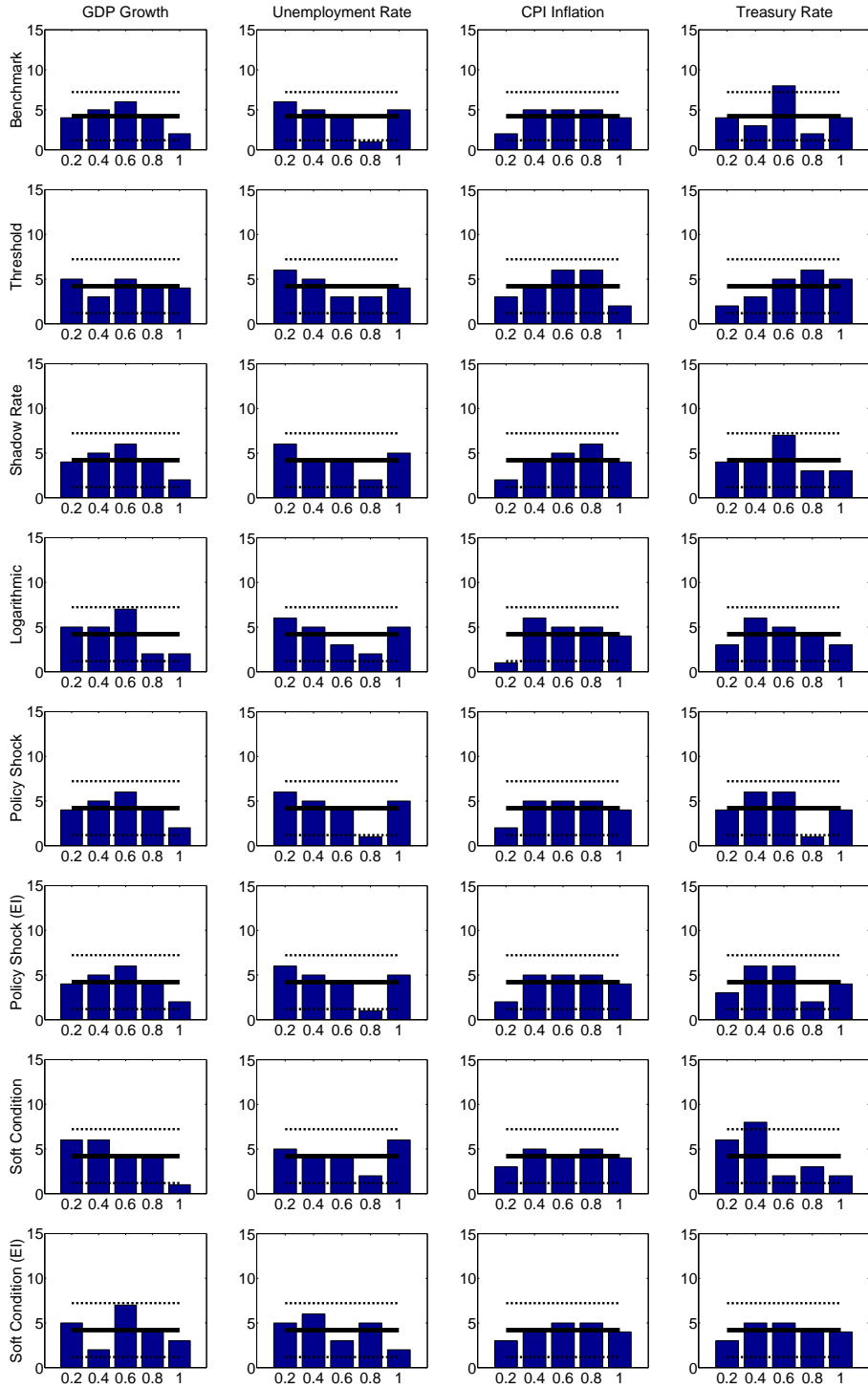
Figure B.2: Probability Integral Transform Histograms ($h = 1$). Notes: this figure shows the probability integral transform histograms for $h = 1$ and the period 2009:1 to 2014:1. The solid line denotes uniformity. The dashed lines are a 90% confidence band. x-axis: quantile; y-axis: number of realizations.

37

Table B.3: Mean Forecast Errors – Difference Specification

	Forecast Horizon h (in Quarters)			
Specification	1	2	3	4
GDP Growth				
Benchmark	0.15	0.12	0.32	0.46
Threshold	0.69	0.67	0.39	0.24
Shadow Rate	0.35	0.35	0.64	0.90*
Logarithmic	0.87*	0.81	1.13	1.60**
Policy Shock	0.15	0.06	0.27	0.42
Policy Shock (EI)	0.15	0.07	0.30	0.44
Soft Condition	0.97**	0.84*	1.10**	1.06**
Soft Condition (EI)	0.10	−0.14	−0.08	0.10
Unemployment Rate				
Benchmark	−0.10	−0.16	−0.20	−0.24
Threshold	−0.10	−0.16	−0.22	−0.27
Shadow Rate	−0.11	−0.17	−0.23	−0.30
Logarithmic	0.02	0.04	0.02	−0.05
Policy Shock	−0.10	−0.17	−0.23	−0.29
Policy Shock (EI)	−0.10	−0.17	−0.22	−0.28
Soft Condition	−0.17	−0.33	−0.50	−0.65
Soft Condition (EI)	−0.11	−0.16	−0.18	−0.18
CPI Inflation				
Benchmark	−0.68	−0.43	−0.10	0.14
Threshold	−0.63	0.20	0.41	0.26
Shadow Rate	−0.83	−0.69	−0.39	−0.11
Logarithmic	−0.40	−0.71	−0.61	−0.22
Policy Shock	−0.69	−0.42	−0.01	0.40
Policy Shock (EI)	−0.69	−0.44	−0.11	0.18
Soft Condition	−0.18	0.10	0.62	1.16***
Soft Condition (EI)	−0.42	−0.26	−0.07	0.19
Treasury Rate				
Benchmark	0.25***	0.52***	0.84***	1.15***
Threshold	0.18	0.31	0.46	0.59
Shadow Rate	0.22***	0.46***	0.75***	1.03***
Logarithmic	0.08	0.16	0.21	0.26
Policy Shock	0.32***	0.63***	0.98***	1.31***
Policy Shock (EI)	0.29***	0.55***	0.85***	1.13***
Soft Condition	0.54***	1.01***	1.41***	1.78***
Soft Condition (EI)	0.28**	0.49**	0.72**	0.98**

Notes: this table shows the mean forecast errors for the period 2009:h to 2014:h. ***, **, and * indicate that the errors are significantly different from zero at the 1%, 5%, and 10% level, respectively.

Table B.4: Root Mean Squared Forecast Errors – Difference Specification

	Forecast Horizon h (in Quarters)			
Specification	1	2	3	4
GDP Growth				
Benchmark	2.55	2.44	2.75	2.18
Threshold	1.17	1.34*	1.39*	1.53
Shadow Rate	1.03	1.03	1.04	1.08
Logarithmic	1.06	1.10	1.17*	1.47**
Policy Shock	1.00	1.02	1.03	1.01
Policy Shock (EI)	1.00	1.03	1.03	1.02*
Soft Condition	1.03	0.96	1.02	1.09
Soft Condition (EI)	0.95	0.89	0.91	0.96
Unemployment Rate				
Benchmark	0.31	0.54	0.70	0.90
Threshold	1.04	1.22***	1.36***	1.49***
Shadow Rate	0.99	0.98*	0.96**	0.96***
Logarithmic	0.85	0.92	1.02	1.11*
Policy Shock	1.00	1.03	1.05	1.06
Policy Shock (EI)	1.00	1.02	1.04	1.05
Soft Condition	1.25	1.34	1.42	1.42
Soft Condition (EI)	1.20	1.25	1.28*	1.27**
CPI Inflation				
Benchmark	2.56	3.14	2.80	1.96
Threshold	1.57	1.38*	1.49	1.34*
Shadow Rate	1.00	0.99	0.96	0.93**
Logarithmic	0.99*	1.07	1.11	1.19*
Policy Shock	1.00	1.01	1.00	0.92
Policy Shock (EI)	1.00	1.01**	1.01	0.89
Soft Condition	0.88	0.92	0.91	1.01
Soft Condition (EI)	0.88	0.90	0.88	0.86
Treasury Rate				
Benchmark	0.48	0.88	1.20	1.52
Threshold	1.14	0.95	0.87	0.82
Shadow Rate	0.96***	0.94***	0.94***	0.93***
Logarithmic	0.91*	0.84*	0.74*	0.69*
Policy Shock	1.09	1.04*	1.04**	1.05
Policy Shock (EI)	1.05	0.97	0.96	0.96
Soft Condition	1.50**	1.41**	1.38**	1.34**
Soft Condition (EI)	1.17	0.99	0.93	0.92

Notes: this table shows the root mean squared forecast errors for the period 2009:h to 2014:h. The errors are provided in levels for the benchmark and as ratios relative to the benchmark for the other specifications. A ratio below unity means that the respective specification outperforms the benchmark. ***, **, and * indicate that errors are significantly different from each other at the 1%, 5%, and 10% level, respectively.

Table B.5: Average Log Predictive Density Scores – Difference Specification

Specification	Forecast Horizon h (in Quarters)			
	1	2	3	4
	GDP Growth			
Benchmark	−2.39	−2.34	−2.43	−2.33
Threshold	−0.24**	−0.36***	−0.41***	−0.32**
Shadow Rate	−0.01	−0.02	−0.01	−0.05**
Logarithmic	−0.06	−0.11	−0.18*	−0.30**
Policy Shock	0.00	−0.00	−0.01	−0.01
Policy Shock (EI)	0.00	−0.00	−0.00	−0.01
Soft Condition	−0.04	−0.02	−0.02	−0.01
Soft Condition (EI)	0.07	0.05	0.04	0.06
	Unemployment Rate			
Benchmark	−0.24	−0.75	−0.96	−1.21
Threshold	−0.02	−0.32**	−0.44***	−0.53***
Shadow Rate	0.03	0.02	0.03	0.04
Logarithmic	0.13	0.00	−0.11	−0.18**
Policy Shock	0.00	−0.03*	−0.04	−0.04
Policy Shock (EI)	0.00	−0.03*	−0.03	−0.04*
Soft Condition	−0.28	−0.40	−0.46	−0.42
Soft Condition (EI)	−0.29	−0.52*	−0.58**	−0.59**
	CPI Inflation			
Benchmark	−2.18	−2.28	−2.15	−2.04
Threshold	−0.34**	−0.27***	−0.23**	−0.30**
Shadow Rate	−0.02	0.01	0.03	0.04**
Logarithmic	−0.05	−0.10	−0.06	−0.08**
Policy Shock	0.00	0.00	0.00	0.00
Policy Shock (EI)	0.00	−0.01	−0.00	0.03
Soft Condition	0.05	0.01	−0.01	−0.05
Soft Condition (EI)	0.06	0.04	0.02	0.05
	Treasury Rate			
Benchmark	−0.71	−1.33	−1.69	−1.96
Threshold	−0.07	−0.06	−0.05	0.08
Shadow Rate	0.04**	0.07***	0.10***	0.10***
Logarithmic	0.09	0.21*	0.37**	0.47**
Policy Shock	−0.04	−0.03	−0.05*	−0.08**
Policy Shock (EI)	0.00	0.04	0.04	0.05
Soft Condition	−0.41***	−0.53***	−0.58***	−0.60***
Soft Condition (EI)	0.00	0.08	0.13	0.13

Notes: this table shows the average log predictive density scores for the period 2009:h to 2014:h. The scores are provided in levels for the benchmark and in differences to the benchmark for the other specifications. A positive difference means that the respective specification outperforms the benchmark. ***, **, and * indicate that scores are significantly different from each other at the 1%, 5%, and 10% level.

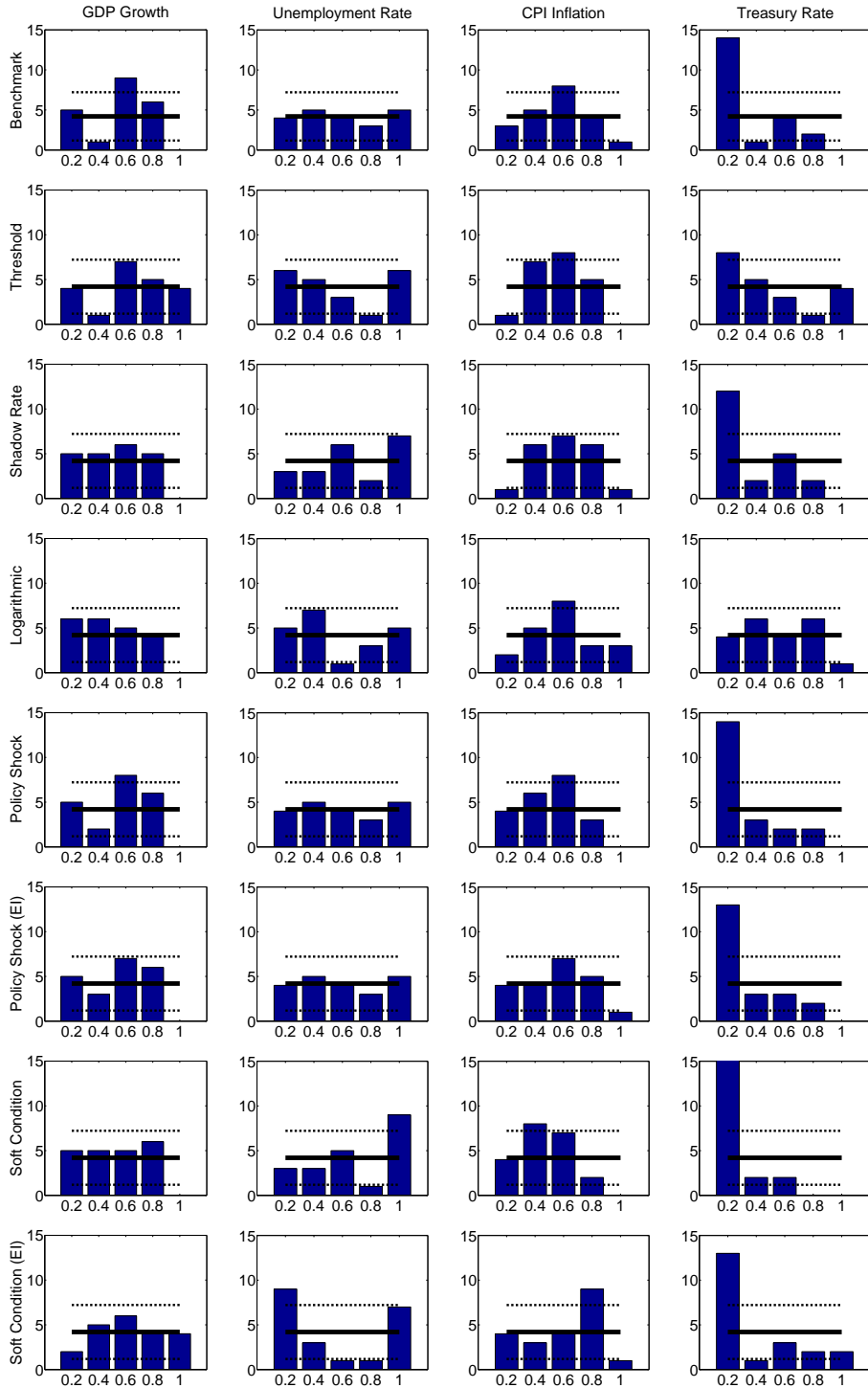Figure B.3: Probability Integral Transform Histograms ($h = 4$) − Difference Specification.
Notes: this figure shows the probability integral transform histograms for $h = 4$ and the period 2009:4 to 2014:4. The solid line denotes uniformity. The dashed lines are a 90% confidence band. x-axis: quantile; y-axis: number of realizations.