Institut für Wirtschaftsforschung Halle
Halle Institute for Economic Research

Qual VAR Revisited: Good Forecast, Bad Story

Makram El-Shagi
Gregor von Schweinitz

December 2012 No. 12

IWH-Diskussionspapiere
IWH Discussion Papers

Authors: *Makram El-Shagi*
Halle Institute for Economic Research
Department of Macroeconomics
E-mail: Makram.El-Shagi@iwh-halle.de
Phone: +49 345 7753 835

Gregor von Schweinitz
Martin-Luther-University Halle-Wittenberg and
Halle Institute for Economic Research
Department of Macroeconomics
E-mail: Gregorvon.Schweinitz@iwh-halle.de
Phone: +49 345 7753 744

Editor:
HALLE INSTITUTE FOR ECONOMIC RESEARCH – IWH
The IWH is a member of the Leibniz Association.

Address: Kleine Maerkerstrasse 8, D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61, D-06017 Halle (Saale), Germany
Phone: +49 345 7753 60
Fax: +49 345 7753 820
Internet: http://www.iwh-halle.de

ISSN 1860-5303 (Print)
ISSN 2194-2188 (Online)

Qual VAR Revisited: Good Forecast, Bad Story

Abstract

Due to the recent financial crisis, the interest in econometric models that allow to incorporate binary variables (such as the occurrence of a crisis) experienced a huge surge. This paper evaluates the performance of the Qual VAR, i.e. a VAR model including a latent variable that governs the behavior of an observable binary variable. While we find that the Qual VAR performs reasonably well in forecasting (outperforming a probit benchmark), there are substantial identification problems. Therefore, when the economic interpretation of the dynamic behavior of the latent variable and the chain of causality matter, the Qual VAR is inadvisable.

Keywords: binary choice model, Gibbs sampling, latent variable, MCMC, method evaluation

JEL Classification: C15, C35, E37

Qual VAR: Gute Prognose, schlechte Geschichte

Zusammenfassung

Aufgrund der aktuellen Finanzkrise verstärkte sich das Interesse an ökonometrischen Modellen, die binäre Variablen (so wie das Auftreten einer Krise) verwenden. Der vorliegende Artikel evaluiert die Leistung des Qual VAR, eines vektorautoregressiven Modells, welches eine latente Variable einbindet, die den Zustand einer binären Variablen bestimmt. Es zeigt sich, dass das Qual VAR eine akzeptable Prognosequalität aufweist (dabei schlägt es im Vergleich ein Probit-Modell). Allerdings existieren substanzielle Identifikationsprobleme. Daraus schließen wir, dass die Verwendung des Qual VAR nicht zu empfehlen ist, wenn die ökonomische Interpretation entweder des dynamischen Verhaltens der latenten Variable oder der Kausalitätsbeziehungen zwischen verschiedenen Variablen von Bedeutung ist.

Schlagwörter: Binary-Choice-Modelle, Gibbs-Sampling, latente Variable, MCMC, Methodenevaluation

JEL-Klassifikation: C15, C35, E37

1 Introduction

Due to the recent financial crisis, the interest in econometric models that allow to incorporate binary variables (such as the occurrence of a crisis) experienced a huge surge. This has led to the development of new models employing qualitative (binary) data, such as artificial neural networks (Fioramanti 2008), binary classification trees (Duttagupta and Cashin 2011) or GARCH-M models with a qualitative variable driving different parameter regimes (Nyberg 2012). At the same time, some older methods developed after the Asian crisis experienced a comeback. These include the signals approach, that has originally been applied to currency and banking crises (Kaminsky and Reinhart 1999). More recent papers use it in an early-warning system for asset price bubbles (Alessi and Detken 2011) or sovereign debt crises (Knedlik and von Schweinitz 2012). Other methods are binary choice such as probit or logit models, developed by Frankel and Rose (1996). They, too, have recently been applied to a variety of crises (Demirgüç-Kunt and Detragiache 2000; Bussière and Fratzscher 2006; Barrell et al. 2010). Furthermore, binary choice models have found their way into research on the dating of business cycle turning points (Rudebusch and Williams 2009; Chauvet and Potter 2010; Nyberg 2010). Those binary choice models assume that an observable binary variable is governed by a latent variable. Harding and Pagan (2011) criticize, that the conventional model cannot capture the time series dependence of the binary variable inherent in current macroeconomic applications. Therefore, Harding and Pagan recommend to use other methods incorporating time series dependence such as the Markov Switching approach. These models have originally been developed for the dating of business cycle turning points (Hamilton 1989; Paap et al. 2009). Like binary choice models, they can also be used for crisis prediction (Fratzscher 2003; Hartmann et al. 2012). Because Markov Switching models estimate the regimes endogeneously, they cannot be applied to predefined binary variables, such as the NBER recessions, IMF interventions and the like. Essentially, while being able to incorporate time series dependence in the binary variable, there is no clear economic definition of what is really meant by the different regimes (El-Shagi et al. 2012).

Some recent methods may overcome both the Harding and Pagan critique and the identification problem of Markov Switching models (Kauppi and Saikkonen 2008; Dueker 2005). Especially the Qual VAR proposed by Dueker (2005) is economically appealing. In the Qual VAR, a latent variable, driving an observable binary variable, and a number of other observables (jointly) follow a VAR process.[1] Estimating a VAR process instead of a single equation, thereby exploiting more information, leads to efficiency gains in the identification of the latent variable. Moreover, since the latent variable can be interpreted as a risk indicator for the event desribed by the binary variable, the VAR structure allows to capture the feedback of the corresponding risk into the economy. The importance of such an interaction between observable and latent variable can for example be observed in the current European debt crisis, where the risk of sovereign default strongly affects government bond interest rates and vice-versa.

A number of recent papers have used the Qual VAR. Bordo et al. (2008) apply the model to bull and bear periods on the stock market, Dueker and Assenmacher-Wesche (2010) assess the recursive forecasting performance of the Qual VAR. This performance is also tested in comparison to other models by Galvão (2006) and Fornari and Lemke (2010). However, the present literature concentrates exclusively on the forecasting performance of the Qual VAR. Furthermore, this evaluation is done based on one specific economic example (i.e. forecasting the 2001 U.S. recession) rather than being performed in a more general framework. It is therefore unclear whether the results are applicable universally. Additionally, while the Qual VAR has been developed for forecasting, the estimates have also been interpreted economically, e.g. by analyzing impulse response functions derived from coefficient estimates and considering the latent variable as in sample measure of event probability (Dueker 2005).

[1] Thus, the Qual VAR is essentially an extension of the dynamic ordered probit of Eichengreen et al. (1985), as observed by Marcellino (2006).

However, even if the forecasting performance was generally high, this is not sufficient to allow a structural interpretation of the results.

Our paper aims at closing these gaps in the literature by providing a range of Monte Carlo studies. This allows a more general examination of the Qual VAR by considering its performance in idealized settings where the data generating process and the latent variable are known. First, we assess the in-sample estimation of the latent variable. Second, we analyze whether or not the Qual VAR identifies the true Granger causality between the latent and other variables. Third, we test the forecasting performance of the estimated system. All these tests are performed for a variety of VAR-specifications covering different chains of causality and uncertainty levels.

The remainder of the paper is organized as follows. In Section 2 we present the estimation technique of the Qual VAR. Section 3 describes the set of Monte Carlo studies used to obtain our results. In Section 4 we discuss identification issues, Section 5 contains the results of our tests of the forecasting performance of the Qual VAR. Section 6 concludes.

2 Estimation of a Qual VAR

The Qual VAR (Dueker 2005) has been developed as a method for forecasting qualitative variables. Originally, it has been applied to the prediction of recessions and business cycle turning points. It assumes, that the present state in the qualitative (usually binary) variable y_t is the observable manifestation of a latent variable y_t^*:

$$y_t = \begin{cases} 0 & , \text{if } y_t^* \leq 0 \\ 1 & , \text{if } y_t^* > 0. \end{cases} \tag{1}$$

The unobservable variable y^* and $k-1$ other observable variables X are said to follow a VAR(p) process:

$$Y_t = \mu + \Phi(L)Y_{t-1} + \epsilon_t, \tag{2}$$

where $Y_t = \begin{pmatrix} X_t \\ y_t^* \end{pmatrix}$, $t = 1, \ldots, T$ is the time index, $\Phi(L)$ is the lag polynomial with the VAR-parameters, μ the constant vector and ϵ_t the error vector at t. Errors are assumed to be multivariate-normally distributed with mean zero and covariance matrix Σ. The covariance matrix, the parameters Φ of the VAR and the unobservable variable y^* are jointly estimated using a Gibbs sampler.

The Gibbs sampler is used to simulate the joint distribution of $\Lambda = (\Phi, \Sigma, y^*)$. In each iteration (i), a value for each element of Λ is randomly drawn from its distribution conditional on the last generated values of all other elements. Depending on the ordering of the elements in the Gibbs sampler, the last generated value can either come from the current or the previous iteration. In the first iteration, starting values for the latent variable are randomly generated (based on the knowledge of the binary variable). Φ and Σ can then be estimated by OLS and used as initial values.

As in standard ML estimation, the coefficients Φ are assumed to be multivariate-normally distributed and the inverse of the covariance matrix of errors Σ^{-1} is assumed to be Wishart distributed. Each y_t^* is drawn from a truncated normal distribution, where the truncation is determined by the observable binary variable y_t.

The order of drawings we use (following Dueker) is $\Phi \rightarrow \Sigma \rightarrow y^*$. That is, we draw $\left(\Phi^{(i+1)} | Y^{(i)}, \Sigma^{(i)}\right)$ and $\left((\Sigma^{-1})^{(i+1)} | \Phi^{(i+1)}, Y^{(i)}\right)$. The vector y^* is sampled element by element, where the distribution of the respective element is conditional on the past of the time series in the current iteration and the future of the time series in the last iteration; i.e., we draw $\left(y_t^{*(i+1)} | \Phi^{(i+1)}, \{y_k^{*(i+1)}\}_{k<t}, \{y_k^{*(i)}\}_{k>t}, X_t, \Sigma^{(i+1)}\right)$. If $p < t < T-p$, $y_t^{*(i+1)}$ is drawn from the exact conditional distribution.[2] It is not feasible to compute these for $t \leq p$ and $t \geq T-p$. We

[2] p is, again, the lag order of the VAR.

follow Dueker (2005), who proposes a Metropolis-Hastings algorithm for the first p periods. For the last p periods, we use simple VAR forecasts to compute the mean of the distribution of $y^{*(i+1)}_{T-p+1}$ and subsequently draw errors from a truncated normal satisfying the conditions imposed by the observable binary variable. Based on $y^{*(i+1)}_{T-p+1}$, we can compute the value of the next period accordingly.

Dueker originally proposed to run the Gibbs sampler once with 10,000 iterations and discard the first 5,000. In contrast, we only use every fifth of the remaining 5,000 iterations to avoid artifacts caused by the dependencies between consecutive iterations in the sampled distributions (Casella and George 1992).[3] From the resulting sample of 1,000 iterations, we calculate median variables Λ_{med}, confidence bands and a set of Fry-Pagan estimate variables Λ_{FP}. Median and confidence bands are calculated for every element in Φ and Σ and y^* separately. However, in the spirit of the Fry-Pagan critique,[4] the set of Fry-Pagan estimates is a consistent set of elements of Λ. Therefore, we select the iteration with the highest joint log likelihood as the Fry-Pagan estimate. The likelihoods are computed assuming multivariate normal distributions for Φ and normal distributions for each element of y^*, where mean and variance are drawn from the distribution obtained from the Gibbs sampler.[5]

3 MCMC setup

Our Monte Carlo study aims to test the capability of Qual VAR to identify the latent variable, to capture the correct chain of causality and to forecast the event driven by the latent variable. To robustly do so, we must test a range of setups, covering the most important features of the data-generating process (DGP) that might affect these issues. The

[3] We deviate from this rule in our forecasting tests as described in Table 2, thereby reducing otherwise exploding runtimes. Estimations show no great difference compared to the other tests.

[4] Fry and Pagan (2007) criticize the use of the pointwise median to construct impulse response functions in SVAR. They argue that while the median is usually the most likely outcome at any point in time, the sequence obtained from the pointwise median values is not itself necessarily a consistent impulse response.

[5] We only use Φ and y for this calculation as the variance of the error of the latent variable is always set to one, making the calculation of a density of the inverse Wishart distribution for Σ impossible.

first important feature of the DGP is the variance of the error term in the equations governing the behavior of the observable variable(s).[6] The variance strongly affects the degree of determination in the system and, thereby, both identification and potential forecasting performances. The second feature of the DGP that we account for is the chain of causality between observable(s) and the latent variable. While this is obviously essential to answer the question whether different chains of causality can be distinguished by Qual VAR estimation, it may also affect identification. Because identification in the Qual VAR exploits both lags and leads, a causality running both directions potentially simplifies correct identification of the latent variable.

We aim to cover all of these aspects in the simplest framework possible. Therefore, the true DGP in all of our simulations uses one observable and the latent variable ($k = 2$) and one lag ($p = 1$). This strongly reduces multicollinearity issues that arise in more complex systems of interacting variables, which may cause trouble in identification. At the same time, such a simple DGP is still capable of capturing a range of potential chains of causality and different degrees of uncertainty. To avoid confusing a lack of power with model uncertainty, the "true" DGP in our simulations exactly mirrors the assumptions of the Qual VAR. That is the event occurs if and only if the latent variable is greater than zero.

Our models are simulated for 200 periods, a sample size that is typically found in macroeconometric time series applications (such as in the original Qual VAR paper by Dueker 2005).[7] The typical economic event that is modeled by the Qual VAR, such as crises, recessions, and business cycle turning points, is rather rare but occurs sufficiently often in the sample period to obtain a certain idea of the underlying dynamics. That is, for our Monte Carlo study, the binary process is required to have multiple (blocks of) events while having

[6] As the scaling of the latent variable is arbitrary, the error variance in its equation is conventionally scaled to one in estimation. Therefore, we do the same in the true process in all of our MCMC experiments.

[7] At the same time, this is a good compromise between short samples that might cause additional small sample problems in the estimation and large samples that rapidly increase computational requirements. We add a swing in phase of 100 periods to each simulation that is dropped before estimation, thereby guaranteeing independence from the starting values.

an overall event probability clearly below 50%. Pretests show that an event probability of 20% satisfies these criteria for all coefficient matrices Φ and covariance matrices of the error terms Σ considered in this paper.

For simplicity, we set the constant term in the observable equation to zero and assume diagonal covariance matrices. Therefore, as the variance of the error term in the latent variable equation, σ_l^2, is held constant at one, the volatility of the system is primarily driven by the variance of the error term in the observable variable equation, σ_o^2. Thus, we can scale this volatility through a single parameter of the MCMC setup. In this context, volatility may have different implications. In forecasting, volatility is a main driver of uncertainty. However, if the observable Granger causes the latent variable, a high σ_o implies that a large share of the volatility of the latent variable can be attributed to changes in the observable variable. This can greatly facilitate the identification of the latent variable. If, on the other hand, the chain of causality implies that the Qual VAR identifies the latent variable mostly through future values of the observable variable (a backward identification), a large σ_o may be an obstacle. To test the influence of σ_o on the identification and forecasting performance of the Qual VAR, we use three different specifications of σ_o: a low volatility specification with $\sigma_o = 0.1$ (labeled *low* in the remainder of the paper), an equal variance specification with $\sigma_o = \sigma_l = 1$ (*eq*), and a high variance specification where $\sigma_o = 10$ (*high*).

We test all three variance specifications in DPGs covering all potential causality chains; i.e., (1) the observable Granger causes the latent variable (labeled *ol* in the remainder of the paper), (2) the latent variable Granger causes the observable variable (*lo*), and (3) the observable and the latent variables mutually Grange cause each other (*olo*). For all model structures covered by our analysis, the matrix Φ is given in Table 1.[8] To construct examples with strong causality chains, the parameter on the off-diagonal is set to 0.7 whenever Granger causality exists. To ensure strong intertemporal dependence and stationarity of the processes

[8]We do not investigate the possibility that the observable and latent variables are independent (with possible autocorrelation).

Table 1: Coefficient matrix Φ for the different causality chains.

		observable $\rightarrow$ latent	
		no	yes
latent $\rightarrow$ observable	no	–	$\begin{pmatrix} 0.9 & 0.7 \\ 0 & 0.2 \end{pmatrix}$
	yes	$\begin{pmatrix} 0.2 & 0 \\ 0.7 & 0.9 \end{pmatrix}$	$\begin{pmatrix} 0.2 & 0.7 \\ 0.7 & 0.2 \end{pmatrix}$

at the same time, the sum of the off-diagonal parameter and the autocorrelation term in the same column is restricted to 0.9.

Given Φ and Σ, the probability of an event, i.e., $P(y^* > 0)$, is driven by the constant term in the latent variable equation. Assuming that the number of events in the sampled process is binomially distributed, we develop an acceptance-rejection algorithm where a VAR simulated with given parameters, covariance matrix and constants is only accepted if the simulated event probability is not statistically different from the 20%. Otherwise, the constant of the latent variable is adjusted and the process is resimulated.

In total, we consider nine different settings of the VAR that differ along two dimensions. In the following, we assess the quality of the in-sample estimation and the forecasting performance of the Qual VAR for all settings. Due to the different requirements of the tests, the number of iterations of the Monte Carlo study and of the Gibbs sampler (applied in each Monte Carlo iteration) is set individually, as outlined in Table 2. For convenience, the Table also lists the specifications of the Gibbs sampler as described in Section 2.

4 Identification Problems

In this section, we analyze whether the Qual VAR is able to correctly identify the latent variable, the parameters and the covariance matrix of the VAR. As described above, the

Table 2: Monte Carlo and Gibbs Sampler setup.

		Identification	Forecasting
	Section	4	5
MCMC	Simulations per model	1'000	10'000
Gibbs Sampler	Total iterations	10'000	4'000
	Swing in iterations	5'000	2'000
	Spacing	5	2
	Final iterations	1'000	1'000

Note: A **spacing** of m means that every m^{th} iteration of the Gibbs sampling is used to compute the final distributions after discarding the first **swing in iterations**. **Final iterations** refers to the number of iterations chosen in that way.

Gibbs sampler produces a distribution of the elements of Λ that we use to calculate the median results and Fry-Pagan estimates $(\Lambda_{med}, \Lambda_{FP})$. Three tests are performed on the estimations of the Qual VAR. First, we determine whether the estimates of the latent variable fit the true latent. A low level of accordance would imply that conclusions drawn from the estimated values of the latent variable must be treated cautiously. We test for unbiasedness using a method from the forecasting evaluation literature (Holden and Peel 1990). However, while a perfect fit is a desirable property, economic conclusions from level differences in the latent variable can be derived if the dynamics are correctly reproduced. Therefore, in our second series of tests, we focus on the explanatory power of the test equation rather than on the coefficient estimates that are usually considered. Because the Gibbs sampler enforces the correct sign of the latent variable, it produces some correlation between the estimated and the true latent variable by construction, even if the economic story behind the latent variable is not correctly captured by the model. Therefore, both tests are performed using non-event periods (i.e., roughly 80% of the sample). The results of those tests are reported in subsection 4.1.

Because our model features considerable persistence, our previous tests may indicate that we correctly capture the dynamics of the system, although the turning points of the latent time series are shifted. In this case, the economic interpretation of the estimated model does not replicate the true data generating process. Therefore, in subsection 4.2, we run a series of Granger causality tests to assess whether the chain of causality implied by the true parameter matrix is correctly identified.[9] A correct estimation of the direction of causality may be enough for a qualitative, although not quantitative, economic interpretation of the results obtained by the Qual VAR.

4.1 Correct estimation of the latent variable

To test if the estimated latent variable is an unbiased estimate of the true latent variable, we regress the two variables.

$$y^* = \alpha + \beta y^*_{true} + \epsilon,^{10} \tag{3}$$

where y^* can be both the median latent variable y^*_{med} or the Fry-Pagan latent y^*_{FP}. In Table 3, we show the estimates of regression (3) as well as the test results for the hypotheses $\alpha = 0$ and $\beta = 1$. As the Monte Carlo simulation is run 1,000 times with different true processes, we present the mean estimate of $(\alpha; \beta)$ and the share of simulations, where t-tests (individually) and F-tests (jointly) reject the two hypotheses.

We find that the hypotheses are rejected in the vast majority of cases in all settings. We generally find $\alpha < 0$ and $\beta < 1$ for both the median and the Fry-Pagan estimate of the bootstrapped distribution. Both estimates are equally unable to reproduce the true latent variable. The small difference in rejection rates can be explained by the fact that the median latent variable essentially is a smoothed version of the Fry-Pagan estimate. The additional noise in y^*_{FP} causes wider confidence bounds around the parameter estimates, which, in turn,

[9] Strictly speaking, our test is more restrictive than a traditional Granger causality test, as our tests require the correct sign of the parameter.

[10] Contrary to the convention in the forecasting literature, our test equation has the true variable as the explanatory variable. This increases readability of the test results. A positive constant now implies a positive bias of the estimates, etc.

Table 3: Estimation of the latent variable and dynamics

	ol, high	*ol, eq*	*ol, low*	*lo, high*	*lo, eq*	*lo, low*	*olo, high*	*olo, eq*	*olo, low*
y^*_{med}	(-0.49;0.17)	(-0.81;0.68)	(-1.10;0.04)	(-1.58;0.31)	(-0.56;0.71)	(-0.24;0.71)	(-0.50;0.27)	(-0.64;0.69)	(-0.19;0.85)
% reject t % reject F	(99.3%;100.0%) 100.0%	(99.9%;94.2%) 100.0%	(100.0%;100.0%) 100.0%	(100.0%;99.7%) 100.0%	(98.2%;94.7%) 99.0%	(99.1%;99.7%) 99.8%	(99.1%;100.0%) 100.0%	(99.9%;96.9%) 99.9%	(99.0%;95.3%) 97.2%
R^2_{med}	0.9834	0.6523	0.0338	0.3385	0.7647	0.9717	0.9562	0.6572	0.9634
y^*_{FP}	(-0.50;0.17)	(-0.84;0.67)	(-1.12;0.03)	(-1.57;0.31)	(-0.53;0.72)	(-0.25;0.71)	(-0.52;0.27)	(-0.65;0.67)	(-0.20;0.84)
% reject t % reject F	(86.8%;100.0%) 100.0%	(98.8%;91.7%) 100.0%	(100.0%;100.0%) 100.0%	(100.0%;99.4%) 100.0%	(92.2%;92.0%) 98.2%	(92.5%;98.8%) 99.4%	(90.7%;100.0%) 100.0%	(99.6%;95.7%) 99.6%	(93.5%;93.3%) 92.1%
R^2_{FP}	0.8617	0.4823	0.0147	0.1705	0.6486	0.9027	0.8361	0.5008	0.8933

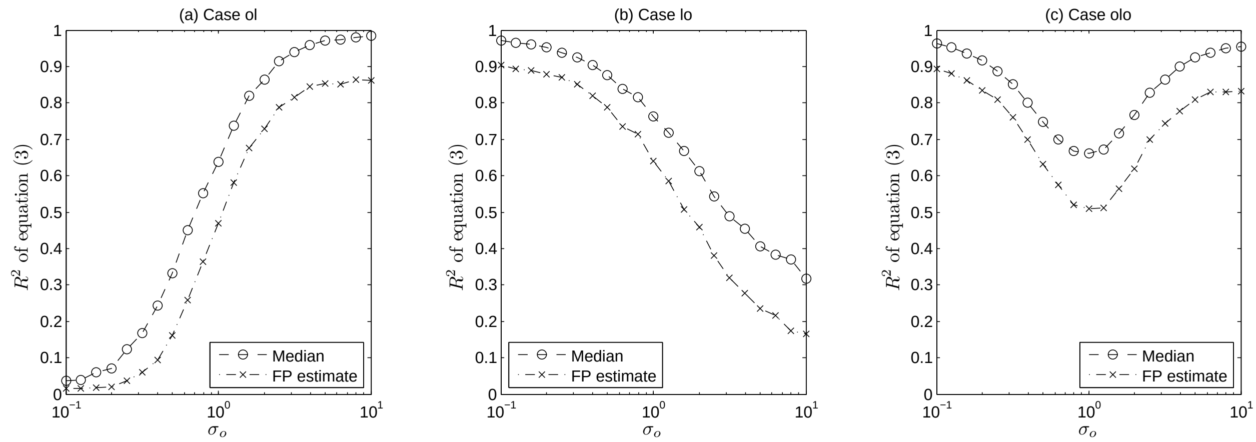
Note: Mean results $(\alpha;\beta)$ of the estimation described in Equation (3) and corresponding t- and F-tests for the median latent variable y^*_{med} and the Fry-Pagan latent variable y^*_{FP}.

lower the rejection rates.

While the dynamics are not completely captured, the mean of the true latent variable in non-event periods is close to the mean of the estimates. Because we exclusively consider negative values of the (true and estimated) latent, $\beta < 1$ implies an upward shift that is compensated by the downward shift implied by $\alpha < 0$. However, although we reject $\beta = 1$ throughout almost all simulations, we often find that $\beta > 0$. In all cases except *ol,low*, β is significantly greater than zero in every bootstrap iteration. That is, the Qual VAR captures at least some of the dynamics of the latent variable in most settings. This is partly reflected by the results of our second test series. However, only in four out of nine cases are the results convincing with an R^2 greater than 0.9. While some more settings (*ol,eq*, *lo,eq* and *olo,eq*) produce at least moderate results with $0.6 < R^2 < 0.8$, it should be considered that the testing environment is rather favorable for the Qual VAR as the true structure of the model (i.e., the selection of variables and the lag order) is known and we consider particularly simple models.

When explaining the Fry-Pagan estimate of the latent with the true latent, R^2 values are even lower (see Figure 1). Again, the reason is the lower degree of noise in the median estimate (compared to the Fry-Pagan estimate). Thus, the difference in R^2 is mostly due to differences in the variance that must be explained, rather than the variance that is explained by the true latent variable.

Whether the Qual VAR captures the dynamics of the latent variable (as described by the R^2) strongly depends on the variance of shocks in the observable equation. If the observable Granger causes the latent variable, estimation is simplified by a high σ_o. If, however, the latent Granger causes the observable variable, a high σ_o strongly decreases the R^2 of our test equation. The reason is that in both cases, the latent variable is mostly identified using information from the observable variable. In the first case (*ol*), the latent variable is strongly correlated to past values of the observable variable. Because the entire variation of the observable variable affects the latent variable, more variance represents information



Note: The reported results are the average of 100 MCMC iterations (instead of the normal 1,000). The reduction was necessary to reduce runtime to a tolerable level.

Figure 1: R^2 of estimation (3) for σ_o between 0.1 and 10, logarithmic equally spaced.

that can be exploited in the estimation. On the contrary, in the second case (*lo*), the latent variable is correlated to future values of the observable variable. However, only the predetermined part of the observable variable contains information on the latent variable. A high σ_o obfuscates the view on the predetermined part of the observable, thus impeding the estimation of y. This difficulty is slightly alleviated (compared to the case *ol*) by the high autocorrelation of the latent variable.

Figure 1 (a) and (b) show the corresponding results for a larger set of different levels of σ_o based on MCMC simulations with fewer iterations (100 instead of 1,000). In the *olo* settings (see Figure 1 (c)), we find a non-monotonic impact of σ_o on R^2. In this case, we can draw information on the latent variable from both past and future values of the observable variable. Initially, when σ_o increases, the loss of information drawn from the future outweighs the gains of information from the past. However, when the variance increases further, the benefit of more information from past observations dominates the impact of σ_o. In our case, with symmetric mutual causality between the latent and the observable variables, the turning point coincides with the equal variance setting (*olo,eq*).

4.2 Correct identification of Granger causality

When testing the correct identification of Granger causality, we report two results. First, for each parameter in Φ, we report the share of iterations where causality is correctly identified (see Table 4). To allow sound economic interpretation, the Qual VAR must capture existing causalities while avoiding the erroneous identification of causalities where none exist. Therefore, if positive true parameters are considered, we report the share of Monte Carlo iterations producing significantly positive estimates of the respective parameter. If parameters that are set to zero are considered, we report the share of iterations producing insignificant results. Second, as a summary of those results, we report the share of iterations where each of the four entries of Φ indicates the correct causality.

We find a substantial share of iterations, where at least one parameter produces an incorrect estimate. Even in the setting where the Qual VAR performs best in this respect, the correct chain of causality is merely identified in less than 70% of the MCMC iterations (see Gr(all) in Table 4). The most frequent reason to reject the joint test is the inability of the Qual VAR to identify weak autoregressive behavior in the latent variable ($\Phi_{2,2} = 0.2$). In three additional settings, there are severe identification problems concerning causality ($\Phi_{1,2}$ and $\Phi_{2,1}$).

First, if the observable Granger causes the latent variable and σ_o is low (*ol,low*), the Qual VAR does not capture this causality in approximately 50% of all iterations. Second, a similarly high rejection rate is found for the opposite case *lo,high*, where we are unable to detect the causality from the latent to the observable. This corresponds to the two settings where the identification of the latent variable is most difficult. Third, with causality running from the latent to the observable variable with low σ_o (*lo,low*), the Qual VAR incorrectly finds a significant effect from the observable on the latent variable in roughly 40% of the Monte Carlo iterations. This is because the actual correlation between past and future values of the observable is much higher in this setting than indicated by the autoregressive

Table 4: Causality and parameter estimates

	ol, high	*ol, eq*	*ol, low*	*lo, high*	*lo, eq*	*lo, low*	*olo, high*	*olo, eq*	*olo, low*
Φ	$\begin{pmatrix} 0.90 & 0.70 \\ 0.00 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.90 & 0.70 \\ 0.00 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.90 & 0.70 \\ 0.00 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.00 \\ 0.70 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.00 \\ 0.70 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.00 \\ 0.70 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.70 \\ 0.70 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.70 \\ 0.70 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.70 \\ 0.70 & 0.20 \end{pmatrix}$
Φ_{med}	$\begin{pmatrix} 0.89 & 0.12 \\ -0.03 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.89 & 0.73 \\ -0.01 & 0.14 \end{pmatrix}$	$\begin{pmatrix} 0.89 & 0.69 \\ -0.00 & 0.15 \end{pmatrix}$	$\begin{pmatrix} 0.19 & 0.00 \\ 0.75 & 0.86 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.05 \\ 0.77 & 0.82 \end{pmatrix}$	$\begin{pmatrix} 0.41 & 0.13 \\ 0.71 & 0.68 \end{pmatrix}$	$\begin{pmatrix} 0.18 & 0.19 \\ 2.42 & 0.19 \end{pmatrix}$	$\begin{pmatrix} 0.18 & 0.67 \\ 0.70 & 0.18 \end{pmatrix}$	$\begin{pmatrix} 0.26 & 0.62 \\ 0.73 & 0.13 \end{pmatrix}$
Φ_{FP}	$\begin{pmatrix} 0.89 & 0.13 \\ -0.02 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.89 & 0.75 \\ -0.01 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.88 & 0.70 \\ 0.00 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.18 & -0.00 \\ 0.97 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.11 & 0.00 \\ 0.91 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.38 & 0.12 \\ 0.77 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.17 & 0.19 \\ 2.49 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.14 & 0.64 \\ 0.80 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0.60 \\ 0.76 & 0.20 \end{pmatrix}$
% t_Φ	$\begin{pmatrix} 91.1\% & 0.0\% \\ 90.5\% & 98.7\% \end{pmatrix}$	$\begin{pmatrix} 90.2\% & 88.8\% \\ 89.3\% & 86.9\% \end{pmatrix}$	$\begin{pmatrix} 87.3\% & 89.9\% \\ 89.4\% & 88.9\% \end{pmatrix}$	$\begin{pmatrix} 90.8\% & 90.0\% \\ 87.4\% & 83.9\% \end{pmatrix}$	$\begin{pmatrix} 87.8\% & 87.0\% \\ 82.9\% & 82.2\% \end{pmatrix}$	$\begin{pmatrix} 9.1\% & 56.3\% \\ 91.1\% & 37.2\% \end{pmatrix}$	$\begin{pmatrix} 88.6\% & 0.0\% \\ 0.2\% & 96.0\% \end{pmatrix}$	$\begin{pmatrix} 87.9\% & 88.5\% \\ 86.6\% & 91.1\% \end{pmatrix}$	$\begin{pmatrix} 62.8\% & 82.8\% \\ 83.4\% & 84.3\% \end{pmatrix}$
% F_Φ	0.0%	69.0%	61.6%	61.4%	62.8%	6.9%	0.0%	66.3%	43.4%
Gr(ind)	$\begin{pmatrix} 100.0\% & 100.0\% \\ 90.5\% & 50.4\% \end{pmatrix}$	$\begin{pmatrix} 100.0\% & 100.0\% \\ 89.3\% & 29.3\% \end{pmatrix}$	$\begin{pmatrix} 100.0\% & 47.1\% \\ 89.4\% & 30.8\% \end{pmatrix}$	$\begin{pmatrix} 80.8\% & 90.0\% \\ 49.1\% & 100.0\% \end{pmatrix}$	$\begin{pmatrix} 71.9\% & 87.0\% \\ 100.0\% & 100.0\% \end{pmatrix}$	$\begin{pmatrix} 99.7\% & 56.3\% \\ 100.0\% & 100.0\% \end{pmatrix}$	$\begin{pmatrix} 76.6\% & 100.0\% \\ 100.0\% & 92.6\% \end{pmatrix}$	$\begin{pmatrix} 74.5\% & 100.0\% \\ 100.0\% & 62.6\% \end{pmatrix}$	$\begin{pmatrix} 99.2\% & 100.0\% \\ 100.0\% & 40.9\% \end{pmatrix}$
Gr(all)	45.3%	26.3%	11.4%	33.6%	61.1%	56.3%	69.6%	38.9%	40.2%
σ_o	10.00	1.00	0.10	10.00	1.00	0.10	10.00	1.00	0.10
$\sigma_{o,med}$	10.02	1.00	0.10	9.97	0.99	0.32	9.78	1.00	0.25
$\sigma_{o,FP}$	9.99	0.99	0.10	9.89	0.92	0.29	9.67	0.94	0.22
% t_{σ_o}	89.6%	90.4%	90.5%	89.6%	88.7%	0.0%	88.2%	89.7%	0.0%

Note: This table contains the true matrix Φ, the median and Fry-Pagan parameter matrix Φ_{med}, Φ_{FP}. Furthermore, it contains the percentage of Monte Carlo iterations, in which the parameter matrix Φ was inside the confidence interval given by the Gibbs sampler (element-by-element in row t_Φ and as a whole in row F_Φ). The same applies to the standard observation of the observable variable, σ_o. Furthermore, we report the share of iterations in which estimated individual parameters captured the true chain of causality individually (Gr(ind)) and jointly (Gr(all)).

parameter of the observable ($\phi_{1,1} = 0.2$). The high persistence of the observable variable is mostly due to the high persistence of the latent variable ($\phi_{2,2} = 0.9$) that is the main driving force of the observable ($\phi_{2,1} = 0.7$ combined with $\sigma_o = 0.1$). Therefore, y^* and correspondingly the observable occurrence of the modeled event (y) is correlated to both past and future values of the observable variable. Accordingly, the autoregressive coefficient of the observable variable is overestimated, as implicit autocorrelation (via the latent) is mistaken for true autoregressive behavior. The autoregressive behavior of the latent variable is underestimated. This is reflected in the parameter estimates. To a lesser extent, the same problem is found with higher values of σ_o for the same causality setting.[11]

5 Forecasting Performance

To assess the forecasting performance of the Qual VAR, we focus on the event probabilities implied by out-of-sample density forecasts of the latent variable of forecast horizons ranging from one to ten periods.

We start with a descriptive analysis of the forecasted probability of the binary event $P(\hat{y}^*_{T+h} > 0)$ in periods where the binary event occurs at the forecast horizon h (i.e. $y_{T+h} = 1$), compared to situations with no event at the forecast horizon (i.e. $y_{T+h} = 0$). This allows us to obtain an impression of the magnitude of absolute forecast errors and their origin. However, this approach has two caveats. First, the occurrence of the binary event in the "far" future is highly uncertain, even when the true data generating process is known. Second, our first test does not compare the performance of the Qual VAR to a reasonably powerful benchmark but merely asks whether the prediction contains any information. Therefore, we run a second series of tests matching Qual VAR forecasts of the conditional probability against (direct) probit forecasts and the unconditional event probability by means of the root mean

[11] As reference, Table 4 also lists the parameter estimates and the share of iterations where the true parameter values are within the estimated confidence bounds. However, the high shares we find for most parameters are often due to extremely broad confidence bounds around the parameters. These broad confidence bounds are caused by the uncertainty concerning the latent variable that is usually poorly identified.

squared error (RMSE). For our application, the RMSE is computed as the difference between the conditional probability implied by the respective model and the conditional probability obtained from the true data generating process for every forecast horizon. Being conditional on information at T, our benchmark accounts for the fact that a certain fraction of the event probability in $T + h$ is actually unforeseeable in T. Because the second test series is concerned with whether the Qual VAR fully exploits available information (rather than its absolute performance), this benchmark is more appropriate than the binary event itself.

Table 5 summarizes the result of our forecast performance evaluation.

5.1 Event prediction

Although we do not formally test against the true probabilities, we report the event probabilities for event and non-event periods derived from the true model as a reference point.[12] In all cases except *ol,low* (where conditional probabilities are always close the unconditional probability of approximately 0.2), the one period ahead conditional probability in event (non-event) periods is above (below) 40% (10%).

In all cases where the true model contains substantial information about the risk of an event in the future, a large share of this is captured by the estimates. Generally, we find that the estimated probabilities are much closer to the probabilities implied by the true model than they are to the unconditional probability. That is, more than half of the risk explained by the true model is also explained by the Qual VAR estimates.

5.2 Forecast comparison

Commonly, RMSEs are computed as the difference between the forecast and the true realization of the variable of interest. Thus, they usually increase over the forecast horizon. On

[12]The number of MCMC iterations performed is sufficient that the difference between event probabilities in event and non-event periods that exists by construction given our AR processes is significantly different from zero.

Table 5: Forecasting with the Qual VAR

		ol, high	*ol, eq*	*ol, low*	*lo, high*	*lo, eq*	*lo, low*	*olo, high*	*olo, eq*	*olo, low*
true forecast, event vs no event, 1,2,5,10 periods		0.96 / 0.00 0.70 / 0.00 0.41 / 0.09 0.27 / 0.17	0.63 / 0.01 0.54 / 0.03 0.36 / 0.10 0.24 / 0.16	0.22 / 0.18 0.20 / 0.18 0.20 / 0.19 0.19 / 0.19	0.65 / 0.01 0.53 / 0.04 0.36 / 0.12 0.26 / 0.18	0.65 / 0.01 0.53 / 0.04 0.35 / 0.12 0.25 / 0.17	0.64 / 0.01 0.52 / 0.04 0.35 / 0.12 0.26 / 0.18	0.93 / 0.00 0.50 / 0.05 0.35 / 0.13 0.26 / 0.19	0.57 / 0.03 0.45 / 0.07 0.32 / 0.14 0.24 / 0.18	0.43 / 0.08 0.42 / 0.09 0.29 / 0.15 0.23 / 0.19
med forecast, event vs no event, 1,2,5,10 periods		0.80 / 0.00 0.61 / 0.01 0.35 / 0.09 0.22 / 0.15	0.62 / 0.01 0.53 / 0.03 0.34 / 0.11 0.23 / 0.17	0.21 / 0.18 0.20 / 0.19 0.20 / 0.19 0.19 / 0.19	0.54 / 0.01 0.40 / 0.05 0.25 / 0.13 0.20 / 0.17	0.54 / 0.02 0.41 / 0.06 0.27 / 0.13 0.20 / 0.17	0.50 / 0.03 0.39 / 0.07 0.27 / 0.14 0.21 / 0.17	0.78 / 0.00 0.44 / 0.06 0.30 / 0.12 0.21 / 0.17	0.55 / 0.03 0.41 / 0.08 0.29 / 0.14 0.22 / 0.18	0.37 / 0.09 0.33 / 0.10 0.25 / 0.16 0.20 / 0.18
FP forecast, event vs no event, 1,2,5,10 periods		0.79 / 0.00 0.61 / 0.01 0.35 / 0.09 0.23 / 0.16	0.62 / 0.01 0.53 / 0.04 0.34 / 0.12 0.24 / 0.17	0.22 / 0.19 0.22 / 0.20 0.21 / 0.20 0.21 / 0.20	0.53 / 0.02 0.41 / 0.07 0.26 / 0.14 0.21 / 0.18	0.54 / 0.02 0.42 / 0.06 0.28 / 0.14 0.22 / 0.18	0.50 / 0.04 0.40 / 0.08 0.27 / 0.14 0.22 / 0.18	0.77 / 0.00 0.44 / 0.06 0.30 / 0.12 0.22 / 0.17	0.54 / 0.04 0.41 / 0.08 0.30 / 0.14 0.23 / 0.18	0.36 / 0.09 0.33 / 0.11 0.25 / 0.16 0.20 / 0.18
RMSE(1)	med FP probit uncond	$0.13^{-/\dagger\dagger\dagger}$ $0.13^{-/\dagger\dagger\dagger}$ 0.10 0.40	$0.06^{***/\dagger\dagger\dagger}$ $0.07^{-/\dagger\dagger\dagger}$ 0.06 0.30	$0.06^{***/\dagger\dagger\dagger}$ $0.08^{-/-}$ 0.07 0.08	$0.12^{***/\dagger\dagger\dagger}$ $0.16^{***/\dagger\dagger\dagger}$ 0.31 0.31	$0.12^{***/\dagger\dagger\dagger}$ $0.15^{***/\dagger\dagger\dagger}$ 0.21 0.31	$0.12^{***/\dagger\dagger\dagger}$ $0.14^{***/\dagger\dagger\dagger}$ 0.17 0.31	$0.11^{***/\dagger\dagger\dagger}$ $0.12^{**/\dagger\dagger\dagger}$ 0.13 0.39	$0.05^{***/\dagger\dagger\dagger}$ $0.07^{-/\dagger\dagger\dagger}$ 0.07 0.28	$0.06^{***/\dagger\dagger\dagger}$ $0.07^{-/\dagger\dagger\dagger}$ 0.07 0.22
RMSE(2)	med FP probit uncond	$0.06^{-/\dagger\dagger\dagger}$ $0.07^{-/\dagger\dagger\dagger}$ 0.03 0.33	$0.04^{-/\dagger\dagger\dagger}$ $0.05^{-/\dagger\dagger\dagger}$ 0.04 0.27	$0.04^{**/\dagger\dagger\dagger}$ $0.06^{-/-}$ 0.05 0.06	$0.13^{***/\dagger\dagger\dagger}$ $0.16^{***/\dagger\dagger\dagger}$ 0.27 0.27	$0.11^{***/\dagger\dagger\dagger}$ $0.14^{***/\dagger\dagger\dagger}$ 0.18 0.27	$0.11^{***/\dagger\dagger\dagger}$ $0.13^{***/\dagger\dagger\dagger}$ 0.14 0.26	$0.05^{***/\dagger\dagger\dagger}$ $0.07^{***/\dagger\dagger\dagger}$ 0.15 0.25	$0.08^{***/\dagger\dagger\dagger}$ $0.11^{***/\dagger\dagger\dagger}$ 0.15 0.23	$0.09^{***/\dagger\dagger\dagger}$ $0.11^{***/\dagger\dagger\dagger}$ 0.14 0.22
RMSE(5)	med FP probit uncond	$0.05^{***/\dagger\dagger\dagger}$ $0.07^{-/\dagger\dagger\dagger}$ 0.06 0.22	$0.04^{***/\dagger\dagger\dagger}$ $0.06^{-/\dagger\dagger\dagger}$ 0.05 0.19	$0.04^{***/\dagger\dagger\dagger}$ $0.05^{-/-}$ 0.05 0.05	$0.12^{***/\dagger\dagger\dagger}$ $0.14^{***/\dagger\dagger\dagger}$ 0.19 0.19	$0.10^{***/\dagger\dagger\dagger}$ $0.12^{***/\dagger\dagger\dagger}$ 0.13 0.19	$0.09^{***/\dagger\dagger\dagger}$ $0.11^{-/\dagger\dagger\dagger}$ 0.11 0.18	$0.05^{***/\dagger\dagger\dagger}$ $0.07^{***/\dagger\dagger\dagger}$ 0.09 0.19	$0.06^{***/\dagger\dagger\dagger}$ $0.08^{***/\dagger\dagger\dagger}$ 0.09 0.16	$0.07^{***/\dagger\dagger\dagger}$ $0.08^{***/\dagger\dagger\dagger}$ 0.09 0.15
RMSE(10)	med FP probit uncond	$0.05^{***/\dagger\dagger\dagger}$ $0.08^{***/\dagger\dagger\dagger}$ 0.08 0.13	$0.04^{***/\dagger\dagger\dagger}$ $0.07^{-/\dagger\dagger\dagger}$ 0.07 0.11	$0.03^{***/\dagger\dagger\dagger}$ $0.05^{-/-}$ 0.05 0.04	$0.09^{***/\dagger\dagger\dagger}$ $0.12^{***/\dagger\dagger\dagger}$ 0.12 0.12	$0.08^{***/\dagger\dagger\dagger}$ $0.10^{***/\dagger\dagger\dagger}$ 0.11 0.12	$0.08^{***/\dagger\dagger\dagger}$ $0.10^{***/\dagger\dagger\dagger}$ 0.11 0.12	$0.05^{***/\dagger\dagger\dagger}$ $0.07^{***/\dagger\dagger\dagger}$ 0.08 0.11	$0.06^{***/\dagger\dagger\dagger}$ $0.08^{-/\dagger\dagger\dagger}$ 0.08 0.10	$0.06^{***/\dagger\dagger\dagger}$ $0.08^{***/\dagger\dagger\dagger}$ 0.08 0.09

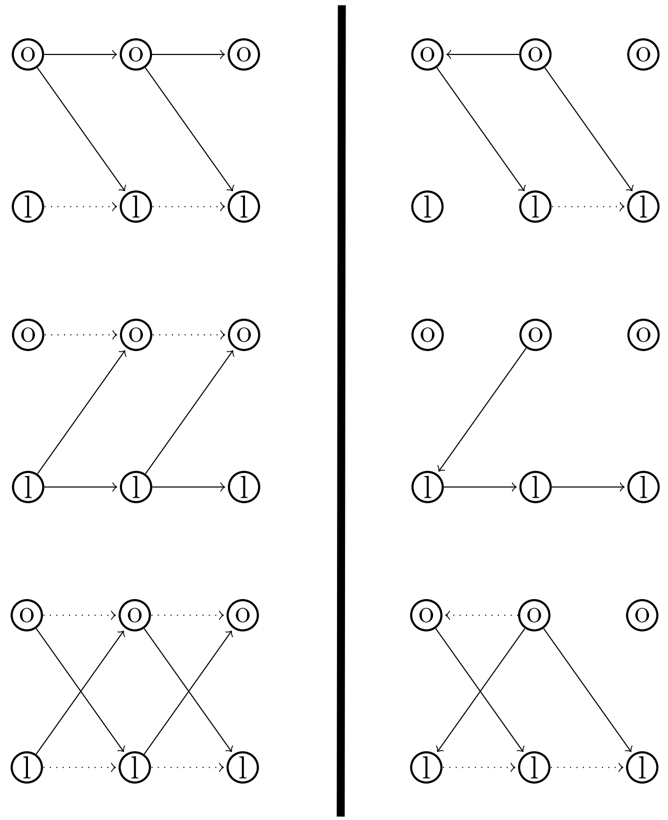
Note: The first three blocks contain the average forecasted probability of an event in event periods vs. non-event periods at forecast horizons of one, two, five, and ten periods. The next four blocks contain, for each of those four forecast horizons, the RMSE of the forecast based on the median latent variable, the Fry-Pagan latent variable, a probit using only the observable variable and that of the unconditional probability. For the first two RMSEs, significant outperformance of the probit at the 1% (5% / 10%) level is indicated by *** (** / *), while significant outperformance of the unconditional probability is indicated by ††† (†† / †).

the contrary, this is not necessarily true for our application, where the RMSE is defined as the root mean squared difference between the forecasted probability and the true conditional probability of an event. Therefore, there are two opposing effects on the magnitude of the RMSE when the forecast horizon increases. First, the forecast uncertainty increases (as with the standard definition of the RMSE). Second, both the true conditional probability and the forecasted probability converge to the unconditional probability. This causes a reduction of the RMSE over the forecast horizon. The dominant effect varies both between different forecast horizons and between settings. Therefore, we find n-shaped, u-shaped and decreasing developments of the RMSE when the forecast horizon increases from one to ten.

In all cases, the median forecast has a lower RMSE than the unconditional probability. The Fry-Pagan estimate performs similarly well, outperforming the unconditional probability in all settings except *ol,low*. In this setting, the noise included in the Fry-Pagan estimate makes it impossible to outperform the unconditional probability, which is – in this case – a quite accurate approximation of the true event probability.

However, the more appropriate benchmark is given by a simple probit model, i.e., a model that exploits the observable information also used in the Qual VAR. Contrary to the probit, the Qual VAR replicates the system dynamics, which are particularly important with respect to forecasting over a long horizon. Therefore, the medium estimate significantly outperforms the probit estimate over the five to ten period horizons in all settings. Again, the Fry-Pagan estimate performs only slightly worse. Whenever the latent variable affects the observable variable (i.e., all *lo* and *olo* settings), these results also hold for the shorter forecast horizons. Only in the *ol* setting with *high* variance in the observable equation can the probit play its strengths and significantly outperform the Qual VAR over short forecast horizons. In *ol,eq*, this holds for a two period ahead forecast because the benefit of explicitly modeling the time series behavior of the latent variable (as accomplished by the Qual VAR) is very limited in the *ol* settings. The latent variable (and thus the event probability) in $t+1$ depends mostly on the observable in t and – to a lesser extent – on the latent variable in t. However, the

true process | **inference channel** ($o_t \to l_{t+1}$)



(a) *ol*

- Strong direct relation between o_t and l_{t+1}
- Strong correlation between o_t and l_t due to common source (o_{t-1})

(b) *lo*

- No direct relation between o_t and l_{t+1}
- Strong correlation between o_t and l_t due to common source (l_{t-1})

(c) *olo*

- Strong direct relation between o_t and l_{t+1}
- Weak correlation between o_t and l_t due to common sources (o_{t-1}, l_{t-1})

Setting/ Timing: t-1 t t+1 | t-1 t t+1

Note: The direction of arrows shows the direct causalities on the left hand side. On the right hand side, arrows point in the opposite direction, if the inference channel from o_t to l_{t+1} uses that detour. Strong links are given by a solid line, weak links are dotted.

Figure 2: Direct causalities in the different settings and inferences (both direct and indirect) from the current observable on the forecasted latent variable.

impact of l_t on l_{t+1} is strongly reflected in the correlation between o_t and l_{t+1} as both o_t and l_t are primarily driven by o_{t-1}. Because the correlation exploited in the probit captures most of the impact of the lagged latent, the value added of the Qual VAR is generally small. If this is combined with situations where the importance of the lagged observable variable (that is included in the probit) is particularly high (e.g. if σ_o is high), the uncertainty carried into the model by trying to identify the dynamic behavior over time overcompensates for the benefits of the identification. This argument is visualized in Figure 5.2 (a).

On the contrary, the probit forecast performs extraordinarily poorly in the *lo* settings. In *lo,high*, probit cannot even outperform an unconditional forecast. Because a causal link between the lagged observable and the contemporary latent – as modeled by the probit – does not even exist in this case, the probit must entirely rely on the correlation between o_t and l_t caused by a common origin (l_{t-1}) (see Figure 5.2 (b)). Especially if σ_o is high, the correlation between o_t and l_t is low, thereby further obfuscating the dynamics.

Finally, in the *olo* cases, there once again is a direct relation between o_t and l_{t+1} (Figure 5.2 (c)). However, as l_{t-1} is only weakly related to l_t (and similarly o_t is only weakly affected by o_{t-1}), the correlation between o_t and l_t is small irrespective of their common sources. Therefore, compared to *ol*, it is more difficult for probit to partly capture the effect of the autoregressive dynamics of the latent variable through the correlation between o_t and l_{t+1}. This explains why Qual VAR can, again, outperform probit regardless of the strong link between the observable and the latent variables.

6 Conclusions

Our results on the performance on the Qual VAR are mixed. The forecasting performance is fairly good. Most notably, compared to a standard procedure in binary forecasting such as a probit, the Qual VAR generally adds substantially. Especially if the dynamic behavior of the latent variable is relevant, the Qual VAR is strong. Even in situations where the Qual

VAR cannot play its strength (such as in short horizon forecasts in our *ol* settings), the loss compared to the probit benchmark is moderate and the absolute forecast errors are minimal.

However, Qual VAR has severe problems in the identification of the economic story. Although the Qual VAR is only confronted with rather simple models in our MCMC framework, it produces substantial errors when estimating the dynamics of the latent variable. Moreover – and at least as problematic from an economic perspective – is the Qual VAR's failure to capture the correct Granger causality in approximately 50% of our simulations. As the Granger causality essentially tells which setting prevails, it is difficult to identify the true setting (i.e., the general economic story). Because the quality of the identification of the latent variable strongly depends on the setting, it is basically impossible to determine whether the Qual VAR results are reliable.

Thus, while providing a good forecasting tool, using the Qual VAR is unadvisable with respect to economic analysis.

Acknowledgments

The authors are highly indebted to Michael Berlemann and Herbert Buscher for valuable comments and suggestions as well as participants of the faculty seminar at the Helmut-Schmidt-University Hamburg.

References

Alessi, L. and Detken, C. (2011), "Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity," *European Journal of Political Economy*, 27, 520–533.

Barrell, R., Davis, E., Karim, D., and Liadze, I. (2010), "Bank Regulation, Property Prices and Early Warning Systems for Banking Crises in OECD Countries," *Journal of Banking & Finance*, 34, 2255–2264.

Bordo, M., Dueker, M., and Wheelock, D. (2008), "Inflation, Monetary Policy and Stock Market Conditions," NBER Working Paper 14019, National Bureau of Economic Research.

Bussière, M. and Fratzscher, M. (2006), "Towards a New Early Warning System of Financial Crises," *Journal of International Money and Finance*, 25, 953–973.

Casella, G. and George, E. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

Chauvet, M. and Potter, S. (2010), "Business Cycle Monitoring with Structural Changes," *International Journal of Forecasting*, 26, 777–793.

Demirgüç-Kunt, A. and Detragiache, E. (2000), "Monitoring Banking Sector Fragility: a Multivariate Logit Approach," *The World Bank Economic Review*, 14, 287.

Dueker, M. and Assenmacher-Wesche, K. (2010), "Forecasting Macro Variables with a Qual VAR Business Cycle Turning Point Index," *Applied Economics*, 42, 2909–2920.

Dueker, M. J. (2005), "Dynamic Forecasts of Qualitative Variables," *Journal of Business and Economic Statistics*, 23, 96–104.

Duttagupta, R. and Cashin, P. (2011), "Anatomy of Banking Crises in Developing and Emerging Market Countries," *Journal of International Money and Finance*, 30, 354–376.

Eichengreen, B., Watson, M., and Grossman, R. (1985), "Bank Rate Policy under the Interwar Gold Standard: a Dynamic Probit Model," *The Economic Journal*, 95, 725–745.

El-Shagi, M., Knedlik, T., and von Schweinitz, G. (2012), "The (Statistical) Significance of the Signals Approach," Discussion Paper 3, IWH.

Fioramanti, M. (2008), "Predicting Sovereign Debt Crises using Artificial Neural Networks: A Comparative Approach," *Journal of Financial Stability*, 4, 149–164.

Fornari, F. and Lemke, W. (2010), "Predicting recession probabilities with financial variables over multiple horizons," Working Paper Series 1255, European Central Bank.

Frankel, J. A. and Rose, A. K. (1996), "Currency crashes in emerging markets: An empirical treatment," *Journal of International Economics*, 41, 351–366.

Fratzscher, M. (2003), "On currency crises and contagion," *International Journal of Finance & Economics*, 8, 109–129.

Fry, R. and Pagan, A. (2007), "Some Issues in Using Sign Restrictions for Identifying Structural VARs," Working Paper 14, NCER.

Galvão, A. (2006), "Structural break threshold VARs for predicting US recessions using the spread," *Journal of Applied Econometrics*, 21, 463–487.

Hamilton, J. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Harding, D. and Pagan, A. (2011), "An econometric analysis of some models for constructed binary time series," *Journal of Business and Economic Statistics*, 29, 86–95.

Hartmann, P., Hubrich, K., Kremer, M., and Tetlow, R. (2012), "Widespread Financial Instabilities and the Macroeconomy-Regime Switching in the Euro Area?" in *mimeo*.

Holden, K. and Peel, D. (1990), "On Testing for Unbiasedness and Efficiency of Forecasts," *The Manchester School*, 58, 120–127.

Kaminsky, G. L. and Reinhart, C. M. (1999), "The twin crises: the causes of banking and balance-of-payments problems," *American Economic Review*, 89, 473–500.

Kauppi, H. and Saikkonen, P. (2008), "Predicting US recessions with dynamic binary response models," *The Review of Economics and Statistics*, 90, 777–791.

Knedlik, T. and von Schweinitz, G. (2012), "Macroeconomic Imbalances as Indicators for Debt Crises in Europe," *Journal of Common Market Studies*, 50, 726–745.

Marcellino, M. (2006), "Leading indicators," *Handbook of Economic Forecasting*, 1, 879–960.

Nyberg, H. (2010), "Dynamic probit models and financial variables in recession forecasting," *Journal of Forecasting*, 29, 215–230.

— (2012), "Risk-Return Tradeoff in US Stock Returns Over the Business Cycle," *Journal of Financial and Quantitative Analysis*, 1, 1–50.

Paap, R., Segers, R., and van Dijk, D. (2009), "Do Leading Indicators Lead Peaks More Than Troughs?" *Journal of Business & Economic Statistics*, 27, 528–543.

Rudebusch, G. and Williams, J. (2009), "Forecasting recessions: the puzzle of the enduring power of the yield curve," *Journal of Business and Economic Statistics*, 27, 492–503.