ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Forecasting Macro Variables with a Qual VAR Business Cycle Turning Point Index

Authors	Michael J. Dueker, and Katrin Wesche
Working Paper Number	2001-019B
Revision Date	October 2001
Citable Link	https://doi.org/10.20955/wp.2001.019
Suggested Citation	Dueker, M.J., Wesche, K., 2001; Forecasting Macro Variables with a Qual VAR Business Cycle Turning Point Index, Federal Reserve Bank of St. Louis Working Paper 2001-019. URL https://doi.org/10.20955/wp.2001.019

Forecasting Macro Variables with a Qual VAR Business Cycle Turning Point Index

February 2005

Michael Dueker*
Federal Reserve Bank of St. Louis
P.O. Box 442, St. Louis, MO 63166
mdueker@stls.frb.org; fax (314) 444-8731

Katrin Wesche
Swiss National Bank
Börsenstrasse 15, CH 8022 Zürich
kas@snb.ch, fax +41-1-631 3901

* The content is the responsibility of the authors and does not represent official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System or the Swiss National Bank.

Forecasting Macro Variables with a Qual VAR Business Cycle Turning Point Index

Abstract

One criticism of VAR forecasting is that macroeconomic variables tend not to behave as linear functions of their own past around business cycle turning points. A large literature focuses therefore on nonlinear forecasting models. This article investigates an alternative linear model that adds to the information set a latent index of nearness to a turning point. The latent index in this VAR is by construction negative during NBER recessions and positive during expansions. The Qual VAR model from Dueker (2005) infers and treats the business cycle turning point index as an additional endogenous variable. We apply the Qual VAR model to recursive out-of-sample forecasting and find that the Qual VAR improves on out-of-sample forecasts from a standard VAR, despite the increase in the number of model parameters that accompanies inclusion of the latent business cycle index.

JEL classifications: **F42, C25, C22**

Key words: **Forecasting, Vector Autoregressions, Business Cycle Turning Points**

For macroeconomic forecasting, the vector autoregression (VAR) has become a standard tool. VARs produce dynamic forecasts that are consistent across equations and forecast horizons. One criticism, however, of VAR forecasting models is that macroeconomic variables do not appear to behave as linear functions of their *own* past around business cycle turning points. One answer has been nonlinear Markov-switching models, which often find that coefficient switches coincide with business cycle turning points [e.g., Hamilton (1989); Filardo (1994); Kim and Nelson (1998); Chauvet (1998)]. Hence, the Markov-switching models tend to confirm the specialness of business cycle turning points. An additional type of nonlinear modeling of macroeconomic data is in Leamer and Potter (2003), who allow a latent business cycle index to alter the dynamics as the threshold variable in a threshold autoregressive model.

As a companion to nonlinear models, however, we propose expanding the information set beyond the own history of the macroeconomic variables by including in VAR forecasts information on business cycle turning points. In this way, the forecasts are conditioned on more than the recent growth rate of output. Suppose, for example, that last month output grew at a 1.5 percent rate but the economy recently had been in a recession, although the economy now appears to have entered a recovery phase. We show that such information on the nearness of the economy to a turning point enhances the out-of-sample forecasts in a VAR model of macroeconomic data. In so doing, we do not claim that the true data-generating process is linear. Instead, we propose that expansion of the information set in linear models by including previously unused qualitative information—such as business cycle turning points—is an avenue worth exploring. In fact, nothing in the way we propose to expand the information set to include turning points prevents application of nonlinear methods—such as Markov switching or threshold behavior.

For this turning point information to be useful in dynamic forecasts of macroeconomic variables, the forecasting model must also produce forecasts of the latent turning point index. Hence we adapt VAR forecasting methods to predict qualitative variables such as business cycle turning points. It is this combination of standard VAR forecasting and a probit-type approach to business cycle turning points that we aim to exploit. Previous methods of forecasting qualitative variables have not resulted in dynamic forecasts; instead, static forecasts are produced from lagged values of the explanatory variables, where the minimum lag length equals the forecast horizon. Our VAR-based model allows leading indicator variables to enter the VAR and to help forecast NBER turning points, as Camacho and Perez-Quiros (2002) aim to do with the leading indicators. Estrella and Mishkin (1998) and Birchenhall et al. (1999), for example, employ a simple probit or logit model to forecast the probability of a business recession in the United States. This paper, in contrast, employs methods for producing dynamic forecasts of a qualitative variable (the NBER recession classifications) within a VAR framework so that the forecasts of turning points are consistent with and enhance the forecasts of the macroeconomic variables. We implement a VAR that adds an additional variable to the vector autoregression—the latent index of nearness to a business cycle turning point. With the additional coefficients that accompany the latent variable, superior out-of-sample forecast performance, relative to the usual VAR model, is by no means assured. In this paper, we assess the out-of-sample forecasting performance of a VAR that includes such a latent variable.

Qual VAR and its Estimation

Our treatment of a qualitative variable within a VAR uses the Qualitative Variable VAR (Qual VAR) from Dueker (2005) and builds on the single-equation dynamic ordered probit model of Eichengreen, Watson and Grossman (1985), Dueker (1999) and Chauvet and Potter (2003). All of these models are designed to permit econometric inference of data that are both qualitative and time series. Dueker (1999) illustrates the relative simplicity of estimating the dynamic probit model through Markov Chain Monte Carlo (MCMC). This extends the work of Albert and Chib (1993) who present MCMC methods for static probit models. MCMC estimation of autoregressive models started with Chib (1993), although Chib worked with autoregressive errors, rather than autoregressive variables as in a VAR. MCMC sampling of an autoregressive latent variable related to a qualitative variable in a VAR was introduced in Dueker (2001, 2005).

To apply MCMC methods to probit-type models, one operationalizes the heuristic notion that a continuous latent variable, y^*, lies behind the observed qualitative variable, $y \in \{0, 1\}$:

$$\begin{aligned} y_t &= 0 \text{ iff } y_t^* \leq 0 \\ y_t &= 1 \text{ iff } y_t^* > 0 \\ y_t^* &= \Psi(L) y_{t-1}^* + \Gamma(L) X_{t-1} + \epsilon_t \qquad (1) \\ & \epsilon_t \sim N(0, 1), \end{aligned}$$

where X_{t-1} is a set of explanatory variables and $\Psi(L)$ and $\Gamma(L)$ are lag polynomials. The qualitative data used for y_t are the recession/expansion classifications determined by the business cycle dating committee at the NBER. In cases like the dynamic probit, where the joint density of the data is difficult to evaluate, data augmentation via MCMC offers a tractable method to generate (i.e., augment the observed data with) a sample of draws from the joint distribution of the y^* through a sequence of draws from the respective conditional distributions. Data augmentation in the present context allows one to treat augmented values of $y_s^*, s \neq t$, as observed data when evaluating the conditional density of y_t^*. Thus, one conditions the density of y_t^* on a *value*, instead of a *density*, of y_{t-1}^*. This argument in favor of MCMC estimation methods applies, and is even amplified, if the latent variable y^* is in a VAR, as opposed to a single equation. Importantly, in a VAR, standard forecasting methods can be applied to predict future y^* and derive dynamic forecasts of all variables in the system.

MCMC Sampling for the Qual VAR

As outlined in Dueker (2005), a Qual VAR model with k variables and p lags is expressed as a standard VAR:

$$\begin{aligned}\Phi(L)Y_t &= \epsilon_t \\ &\epsilon_t \sim \text{Normal}(0, \Sigma)\end{aligned} \tag{2}$$

where Y_t is a $k \times 1$ vector consisting of macroeconomic data, X_t, plus the latent business cycle turning point index y^*; $\Phi(L)$ is a set of $k \times k$ matrices, from $L = 0, .., p$, with the identity matrix at $L = 0$. Hence, the VAR regression coefficients are in $\Phi(L)$. The parameters that require conditional distributions for MCMC estimation are Φ, y^* and Σ, the covariance matrix.

MCMC is an attractive estimation procedure for the Qual VAR because the conditional distribution of the latent variable is straightforward given the VAR coefficients; in turn, the conditional distributions of the VAR coefficients come from a Bayesian VAR, given values for the latent variable. (Discussion of the sampling distribution of the latent turning point index is in the appendix; the estimation of the Bayesian VAR uses code generously supplied by John Robertson and Ellis Tallman, which implements directly the Bayesian VAR proposed by Sims and Zha (1998).) The key idea behind MCMC estimation is that after a sufficient number of iterations, the draws from the respective conditional distributions jointly represent a draw from the joint posterior distribution, which often cannot be evaluated directly [Gelfand and Smith (1990)].

Markov Chain Monte Carlo estimation of this model consists of a sequence of draws from the following conditional distributions, where superscripts indicate the iteration number:

$$\begin{array}{rcl}
\text{VAR coefficients} & \sim & \text{Normal} \\
f(\Phi^{(i+1)} & | & \{y_t^{*(i)}\}_{t=1,..,T}, \{X_t\}_{t=1,..,T}, \Sigma^{(i)}) \\
\text{Covariance matrix} & \sim & \text{inverted Wishart} \\
f(\Sigma^{(i+1)} & | & \{y_t^{*(i)}\}_{t=1,..,T}, \{X_t\}_{t=1,..,T}, \Phi^{(i+1)}) \\
\text{latent variable} & \sim & \text{truncated Normal} \\
f(y_t^{*(i+1)} & | & \Phi^{(i+1)}, \Sigma^{(i+1)}\{y_j^{*(i+1)}\}_{j<t}, \{y_k^{*(i)}\}_{k>t}, \{X_t\}_{t=1,..,T}).
\end{array} \quad (3)$$

Conditional on a set of values for y^*, the VAR coefficients, Φ, are normally distributed. The covariance matrix is part of a normal-inverted Wishart conjugate pair with the VAR coefficients. The inverted Wishart conditional distribution of Σ is presented in Chib and Greenberg (1996). Further details on these conditional distributions and prior distributions are in the appendix.

For the in-sample observations, each observation of the latent variable, y^*, has a truncated normal distribution, where y^* is not allowed to be negative during expansions or positive during recessions. Using the conditional mean, μ_{y^*}, and variance, σ_{y^*}, derived in the appendix, $(y^* - \mu_{y^*})/\sigma_{y^*}$ is in the interval $(-\infty, 0 - \mu_{y^*}/\sigma_{y^*}]$ during recessions and $(0 - \mu_{y^*}/\sigma_{y^*}, +\infty)$ during expansions. Let the relevant bounds be denoted (l, u) and F be the cumulative normal density function. To sample from the truncated normal, we first draw a uniform variable, υ, from the interval $(F(l), F(u))$. The truncated normal draw for $(y^* - \mu_{y^*})/\sigma_{y^*}$ is then $F^{-1}(\upsilon)$.

Data

The Qual VAR uses monthly data for the United States from January 1959 to December 2003. The VAR is specified with 6 lags of each variable. The qualitative variable is a binary 0/1 variable that denotes recessions and expansions, with switches at business cycle turning points. The recession and expansion classification comes from the NBER. The macroeconomic variables, X, in the VAR include industrial production, the annualized month-to-month change in the consumer price index (CPI), the Fed funds rate, and the spread between the 10-year government bond yield and the 3-month treasury bill rate. All data are from Federal Reserve Bank of St. Louis economic database (FRED).

Variables such as these have been widely used in the literature to forecast recessions [e.g. Estrella and Mishkin (1998)]. Industrial production reflects output in the manufacturing sector, which is closely related to the business cycle. Sometimes the cycle itself is defined in terms of industrial production, although the NBER methodology stresses that in a recession many economic time series show contraction. Inflation generally is positively correlated with real activity. Financial variables are also often used to predict the business cycle. The short-term interest rate is included because most central banks use it as the monetary policy instrument. Romer and Romer (1989) claim that most US recessions are preceded by an exogenous tightening of monetary policy. In many studies, the yield-curve spread has been found to help to predict the business cycle [Estrella and Mishkin (1998); Bernard and Gerlach (1998)].

It is well known that VARs often perform poorly in forecasting as the number of unknown coefficients is large relative to the amount of information in the data. One solution is to reduce parameter uncertainty by placing restrictions on the coefficients in the form of Bayesian prior information. The prior distribution for the VAR coefficients follows Sims and Zha (1998). We use an informative prior for the coefficient matrix Φ and a diffuse prior for the covariance matrix, Σ. The restrictions on Φ follow the idea of Litterman (1986) that beliefs are centered on a random walk model. The hyperparameters to control the priors on the Bayesian VAR are taken from Robertson and Tallman (2001). They report values of the hyperparameters that performed well in a forcasting VAR model for the US economy.[1] Both the Qual VAR and the ordinary VAR are estimated with the same priors by Gibbs sampling.

Expanding the information set generally results in a better fit in-sample. Due to the problem of overfitting it is not clear, however, if also the out-of-sample forecasting performance is improved. We therefore evaluate the forecasting performance of the Qual VAR model relative to a simple model by comparing the out-of-sample forecasts. For out-of-sample forecasting the last 4 years of data are used, i.e., the exercise starts with January 2000. For the first forecast, the Qual VAR thus is estimated with data up to December 1999 and out-of-sample forecasts at the 1, 3, 6, and 12 month horizon are computed. Thereafter, the data set is extended by one observation, the Qual VAR is reestimated and a new set of forecasts is obtained. For each estimation, we use 1200 iterations of the Gibbs sampler after having discarded the first 300 draws to let the Gibbs sampler converge to the posterior distribution. To obtain forecasts from the simple VAR, the same procedure is followed. At each point in time we thus use only the information that would have been available for the forecaster at that time. Note that nevertheless these are not real-time forecasts as we use revised data as of mid 2004 and not the original real-time data. Another problem with real-time forecasting using a business cycle indicator is the considerable time lag with which the NBER releases information that a turning point has occured. For example, the NBER declared only on July 17, 2003 that the last trough had occured in November 2001.

[1] The hyperparameters are defined in Robertson and Tallman (2001) and we set the parameters controlling the priors to the values suggested by Robertson and Tallman (2001): $\lambda_0 = 0.6$, $\lambda_1 = 0.1$, $\lambda_2 = 1$, $\lambda_3 = 0.1$, $\mu_1 = 5$, and $\mu_2 = 5$.

Recursive Out-of-Sample Forecasting

The distance of the latent variable, y^*, from zero indicates how far the economy is from a business cycle turning point; that is, it shows how large a shock would have to be for the latent variable to reach zero. Figure 1 shows the posterior mean of the latent variable for the full sample period, together with the NBER recession dates. The posterior mean of the latent variable incorporates the information from the macroeconomic variables in the VAR and can be regarded as a business cycle index, measuring the strenghts of expansions and recessions. By construction the business cycle index is forced below zero at business cycle turning points.

For the augmented y^* to be a useful addition to the dynamic forecasts, the VAR must be able to forecast y^* accurately because dynamic forecasts build forecasts from forecasts at shorter horizons. For this reason, we present forecast results for y^* itself, in addition to results on the forecasted probability of recession. Unlike the former, the latter cannot be interpreted as a forecast of the distance from a turning point, but it is the forecasting measure available from simple probit models for the purposes of comparison.

Forecasts of the Latent Variable

Figure 2 shows the posterior mean of the latent variable y^*, together with the 1, 3, 6 and 12 months-ahead out-of-sample forecasts from January 2000 on. The shaded area indicates the 2001 recession. Although the latent variable is forced below zero when a recession occurs, forecasts of it do not have to cross the zero line at business cycle turning points. At forecast horizons of one and three months, forecasts track the latent variable across the zero line. The onset of the recession is predicted with a lag of 1 and 3 months respectively. At the 6 month horizon, the latent variable shows a marked dip during the recession but the zero line is not reached. At a forecasting horizon of 12 months, the forecast flattens and the recession is not foreseen that far ahead.

All in all, the forecasting results for y^* suggest that the Qual VAR forecasts the latent variable sufficiently well for forecasts of y^* to be a useful addition to the forecast information set used to produce dynamic forecasts of the other macroeconomic variables.

Recession-Probability Forecasts

Simple probit models measure their forecasting performance in terms of the forecasted probability of recession, so we computed this measure for the Qual VAR and a simple probit. For the simple probit model, we estimate the model with the explanatory variables, X, lagged k periods and obtain the probabilities from the cumulative density function used to evaluate the likelihood function:

$$Pr(R_t = 1 \mid I_{t-k}) = \Phi(\beta_0 + \beta_1 X_{t-k}),$$

with Φ denoting the cumulative standard normal density function, $R_t = 1$ meaning that the economy is in recession in period t and X_{t-k} being the set of explanatory variables. Since we assume in the Qual VAR that the qualitative variable is known, we allow also the lagged value of the indicator to enter the simple probit. As recessions do not occur often, the lagged value of the dependent variable is likely to improve forecasting performance. As a starting point, we also include 6 lags of the explanatory variables in the simple probit. Since the model is likely to be overparameterized with such a large number of lags, we subsequently eliminate insignificant variables until the Schwarz information criterion is minimized. This procedure is performed separately for each forecasting horizon so that the final specification of the probit model for each forecasting horizon differs.

To arrive at the forecasted recession probability using the Qual VAR, we simulate the Qual VAR system k periods from the forecast date and calculate the forecasted recession probability as the percentage of iterations where the simulated latent variable was negative k periods into the future. This simulation uses the covariance matrix of the VAR innovations, Σ, so that the simulation only requires draws of independent standard normals, ϵ,

$$\hat{Y}_{t+k} = E[\hat{Y}_{t+k} \mid \hat{Y}_{t+k-1}] + \Sigma^{\frac{1}{2}} \epsilon_{t+k},$$

where $\hat{Y}$ denotes a simulated value. If the simulated value of the latent variable, $\hat{y}^*$, which is the last element in $\hat{Y}$, is less than zero, for example, in 30 percent of the simulation iterations, then the forecasted probability of recession is 30 percent.

Table 1 reports posterior means of the root mean squared error (RMSE) and classification errors for forecasts from the Qual VAR and compares them to forecasts from the simple probit. Except for the 12-month forecast horizon, the Qual VAR model has a lower RMSE than the probit model. A similar picture emerges if one looks at the classification of the forecasted probabilities into recession and expansion. We use a threshold of 0.5 to separate recessions from expansions.[2] The table shows the number of wrong predictions using this threshold. At short forecasting horizons of 1 and 3 months the classification errors from the Qual VAR model are markedly lower than those from the simple probit.

Forecasts of the Macroeconomic Variables

Having established that the Qual VAR produces servicable forecasts of business cycle turning points and the nearness to a turning point, we now turn to the key question: Does inclusion of the turning point information enhance the forecasts of output in an augmented Qual VAR system, relative to an ordinary VAR consisting of a conventional set of macroeconomic variables? In other words, does the appended turning point measure help overcome the

[2] Results remain unchanged with other thresholds.

criticism that macroeconomic variables do not behave as linear functions of their *own* past around business cycle turning points, as assumed by standard VARs? Using the Qual VAR to produce dynamic forecasts of all variables in the system, including the turning point measure, y^*, we concentrate on the forecasting fit for industrial production, the most widely used measure of output at the monthly frequency. Though the models include industrial production in levels, forecast statistics are reported for the monthly growth rate of industrial production.

Table 2 compares the posterior mean of the RMSE from the QualVAR with the posterior mean of the RMSE from the simple VAR without the qualitative variable. For all variables, RMSEs rise with the forecasting horizon, but most significantly so for the interest rate and the term spread. Except for inflation, the RMSE from the QualVAR is always lower than the corresponding RMSE from a simple VAR. This is all the more remarkable as this concerns out-of-sample forecasts.

Table 3 gives the difference between the RMSEs for the Qual VAR and an ordinary VAR in percent, averaged over the runs of the Gibbs sampler. A positive value means that the RMSE for the simple VAR exceeds the RMSE for the Qual VAR. Except for inflation at the 6 and 12 month horizon, the RMSE from the Qual VAR is always lower than the RMSE from the simple VAR. The improvement in forecasting performance is most apparent for industrial production. Up to the six-month horizon, the RMSEs for industrial production growth are about 12 percent higher for the ordinary VAR. Even at a horizon of 12 months, the RMSE from the ordinary VAR is about 5 percent higher than those from the Qual VAR. But also forecasts for the interest rate and the term spread are improved by up to 17 percent, especially at long horizons from 6 to 12 months.

The second entry in Table 3 gives the percentage of draws that is less than zero, which indicates the probability of an improvement in forecasting precision. For both models, RMSEs and their differences are computed at each iteration for each variable at the 1, 3, 6, and 12 month horizon. Gibbs sampling allows us to compute forecast RMSEs at each iteration. We can thereby obtain "empirical" probability levels for the difference in the RMSE between the simple VAR and the Qual VAR by counting the number of iterations in which this difference is negative. For industrial production, the interest rate, and the term spread the improvement in forecasting ability is—with a single exception—uniformly better for the Qual VAR in more than 95 percent of the draws. That is, when we calculate percentile bands on the forecast RMSEs, the RMSEs for the simple VAR would generally lie above the Qual VAR's 90 percent band. Only for inflation is the forecasting record mixed between the Qual and simple VARs.

The robustness of this forecast improvement from the Qual VAR across all forecast horizons adds credence to the claim that turning points can be a crucial piece of information to accurate forecasts. Hence, the Qual VAR has improved the forecast performance, not by adding nonlinearitites to a fixed-dimension system, but by expanding the dimension of the system of equations to apply information from a relevant qualitative variable—the recession/expansion classification that is delineated by business cycle turning points.

Conclusion

Despite the ease with which VAR forecasting models produce dynamic multi-period forecasts, a valid criticism of VARs is that macroeconomic variables tend not to behave as linear functions of their own past around business cycle turning points. A common response has been to add nonlinearities to forecasting models, especially in the form of Markov-switching coefficients. One drawback of this approach is that nonlinear forecasting models do not produce dynamic forecasts as readily as VARs.

Here we propose a different response that does not view linearity as the culprit; instead, we allow macroeconomic variables, such as industrial production, to be functions not only of their own past values but other relevant conditioning information as well. In this case, the relevant information happens to involve qualitative data on business cycle turning points. Therefore, we follow Dueker (2005) and extend VAR analysis to include dynamic forecasting of qualitative variables—a Qual VAR model. The Qual VAR model is estimated using Markov Chain Monte Carlo methods in order to sample from the posterior distribution of the continuous latent variable that lies behind the qualitative data. Conditional on values for the latent variable, standard VAR forecasting projections lead to dynamic forecasts.

We apply the Qual VAR model to US data and find that the Qual VAR improves on forecasts from a standard VAR model. The Qual VAR approach to forecasting industrial production leads to improved forecasts, relative to an ordinary VAR, at all horizons through 12 months, but especially at shorter horizons through six months.

In sum, our approach investigates the merits of linear forecasts based on an expanded forecast information set that includes information regarding qualitative variables, such as business cycle turning points. Previous work investigated the merits of nonlinear forcasts based only on the past values of the macroeconomic variables themselves. Subsequent research can work towards melding these two potentially complementary approaches.

References

Albert, J. H. and S. Chib, (1993): “Bayesian Analysis of Binary and Polychotomous Response Data,” *Journal of the American Statistical Association* 88 (422), pp. 669-679.

Bernard, H. and S. Gerlach, (1998): “Does the Term Structure Predict Recessions? The International Evidence,” *International Journal of Finance and Economics* 3 (3), pp. 195-215.

Birchenhall, C. R., H. Jessen, D. R. Osborn and P. W. Simpson, (1999): “Predicting US Business Cycle Regimes,” *Journal of Business and Economic Statistics* 17 (3), pp. 313-323.

Camacho, M. and G. Perez-Quiros, (2002): This is What the Leading Indicators Lead,” *Journal of Applied Econometrics* 17 pp. 61-80.

Chauvet, M., (1998): An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching,” *International Economic Review* 39 (4), pp. 969-996.

Chauvet, M. and S. Potter, (2003): “Predicting Recessions using the Yield Curve,” *Journal of Forecasting* forthcoming.

Chib, S., (1993): “Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach,” *Journal of Econometrics* 58 pp. 275-294.

Chib, S. and E. Greenberg, (1996): “Markov Chain Monte Carlo Simulation Methods in Econometrics,” *Econometric Theory* 12 pp. 409-431.

Dueker, M. J., (1999): “Conditional Heteroscedasticity in Qualitative Response Models of Time Series: A Gibbs Sampling Approach to the Bank Prime Rate,” *Journal of Business and Economic Statistics* 17 (4), pp. 466-472.

Dueker, M. J., (2001): Forecasting Qualitative Variables with Vector Autoregressions: A Qual VAR Model of U.S. Recessions,” Federal Reserve Bank of St. Louis, *Working Paper* No. 012A.

Dueker, M. J., (2005): “Forecasting Qualitative Variables with Vector Autoregressions: A Qual VAR Model of U.S. Recessions,” *Journal of Business and Economic Statistics* forthcoming.

Eichengreen, B. J., M. W. Watson and R. S. Grossman, (1985): "Bank Rate Policy Under the Interwar Gold Standard: A Dynamic Probit Model," *Economic Journal* 95 pp. 725-745.

Estrella, A. and F. S. Mishkin, (1998): "Predicting U.S. Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics* 80 (1), pp. 45-61.

Filardo, A. J., (1994): "Business-Cycle Phases and Their Transitional Dynamics," *Journal of Business and Economic Statistics* 12 (3), pp. 299-308.

Gelfand, A. E. and A. F. M. Smith, (1990): "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* 85 (410), pp. 398-409.

Hamilton, J. D., (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57 (2), pp. 357-384.

Kim, C.-J. and C. R. Nelson, (1998): "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching," *Review of Economics and Statistics* 80 (3), pp. 188-201.

Leamer, E. and S. Potter, (2003): "A Nonlinear Model of the Business Cycle," manuscript, Federal Reserve Bank of New York.

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

Robertson, J. C. and E. W. Tallman, (2001): "Improving Federal-Funds Rate Forecasts in VAR Models Used for Policy Analysis," *Journal of Business and Economic Statistics* 19 (3), pp. 324-330.

Romer, C. D. and D. H. Romer, (1989): " Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," **O. J. Blanchard and S. Fischer** (ed.) *NBER Macroeconomics Annual*, MIT Press, Cambridge, pp. 121-170.

Sims, C.A. and T.A. Zha, (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review* 39 (4), pp. 949-968.

Appendix: Qual VAR Estimation

Conditional distribution of the latent variable

Here we derive the mean and variance of the underlying normal distribution for y^*. Taking Y as the set of data that follows a vector AR(p) process, we need to derive the conditional distribution of $y_{it} \mid Y_{-t}, Y_{t/i}$, where Y_{-t} is the full vector time series except for time t data and X_t is the vector at time t except for the latent variable y_t^*.

We start by deriving the conditional distribution of $Y_t \mid Y_{-t}$. Because the autoregressive order is p, Y_t will affect the residuals for $p+1$ periods:

$$\begin{aligned}
\epsilon_t &= Y_t - \mu - \phi_1 Y_{t-1} - ... - \phi_p Y_{t-p} \\
\epsilon_{t+1} &= Y_{t+1} - \mu - \phi_1 Y_t - ... - \phi_p Y_{t-p+1} \\
\vdots \quad & \quad \vdots \\
\epsilon_{t+p} &= Y_{t+p} - \mu - \phi_1 Y_{t+p-1} - ... - \phi_p Y_t
\end{aligned} \tag{4}$$

Let κ_j be the known part of ϵ_j under the assumption that the values of the latent variable from other time periods are taken as given:

$$\begin{aligned}
\epsilon_t &= Y_t + \kappa_t \\
\epsilon_{t+1} &= \kappa_{t+1} - \phi_1 Y_t \\
\vdots \quad & \quad \vdots \\
\epsilon_{t+p} &= \kappa_{t+p} - \phi_p Y_t
\end{aligned} \tag{5}$$

The density of $(\epsilon_t, ..., \epsilon_{t+p})$ can be written as a function of Y_t:

$$\begin{aligned}
-\tfrac{1}{2} & \quad (Y_t + \kappa_t)' \Sigma^{-1} (Y_t + \kappa_t) \\
-\tfrac{1}{2} & \quad (\kappa_{t+1} - \phi_1 Y_t)' \Sigma^{-1} (\kappa_{t+1} - \phi_1 Y_t) \\
& \quad \vdots \\
-\tfrac{1}{2} & \quad (\kappa_{t+p} - \phi_p Y_t)' \Sigma^{-1} (\kappa_{t+p} - \phi_p Y_t),
\end{aligned} \tag{6}$$

where Σ is the cross-equation covariance matrix of the errors, which are uncorrelated across time.

After collecting all cross-products, we have

$$Y_t \mid Y_{-t} \sim N(C^{-1}D, C^{-1}), \tag{7}$$

where

$$C = (\Sigma^{-1} + \phi_1'\Sigma^{-1}\phi_1 + ... + \phi_p'\Sigma^{-1}\phi_p)$$

and

$$D = (-\Sigma^{-1}\kappa_t + \phi_1'\Sigma^{-1}\kappa_{t+1} + ... + \phi_p'\Sigma^{-1}\kappa_{t+p}).$$

Thus, it is convenient to define $\tilde{Y}_t = Y_t - C^{-1}D$, where

$$f(\tilde{Y}_t) \quad \propto \quad exp\{-.5\tilde{Y}_t'C\tilde{Y}_t\}$$

Without loss of generality, we can assume that y_t^* is the last element in Y_t and partition C accordingly:

$$C \quad = \quad \begin{pmatrix} C_{00} & C_{01} \\ C_{01}' & C_{11} \end{pmatrix}.$$

After collecting terms, we have

$$\tilde{y}_t^* \mid \tilde{X}_t \sim N(-C_{11}^{-1}C_{01}\tilde{X}_t, C_{11}^{-1}). \tag{8}$$

The conditional mean of y_t^* is the conditional mean of $\tilde{y}_t^*$ plus the bottom right-hand element of $C^{-1}D$.

Figures and Tables

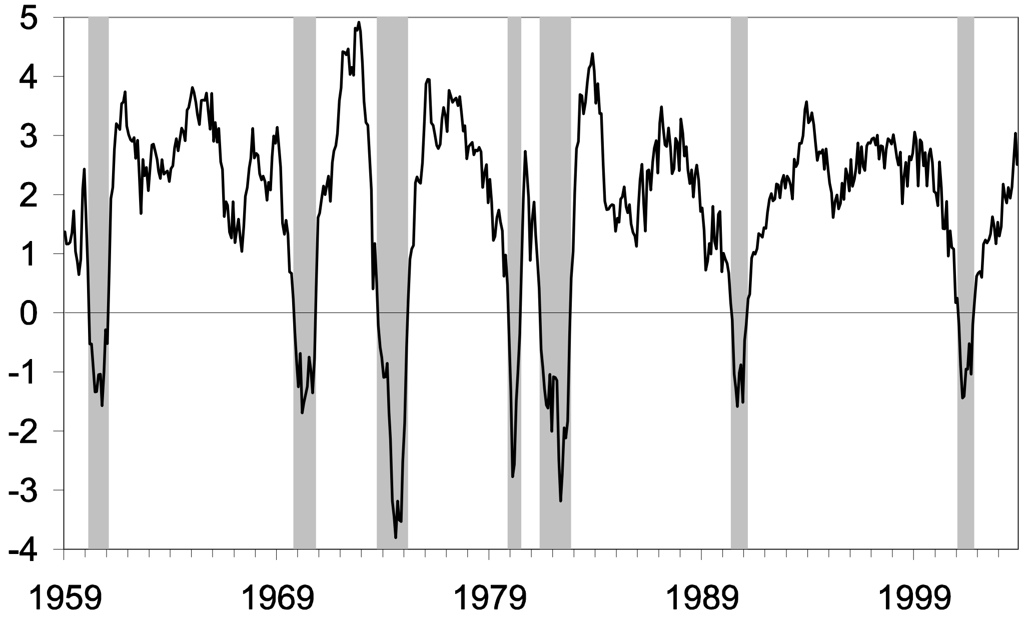

Figure 1: *Posterior Mean of Latent Variable.* Shaded areas indicate recession dates from the NBER classification.

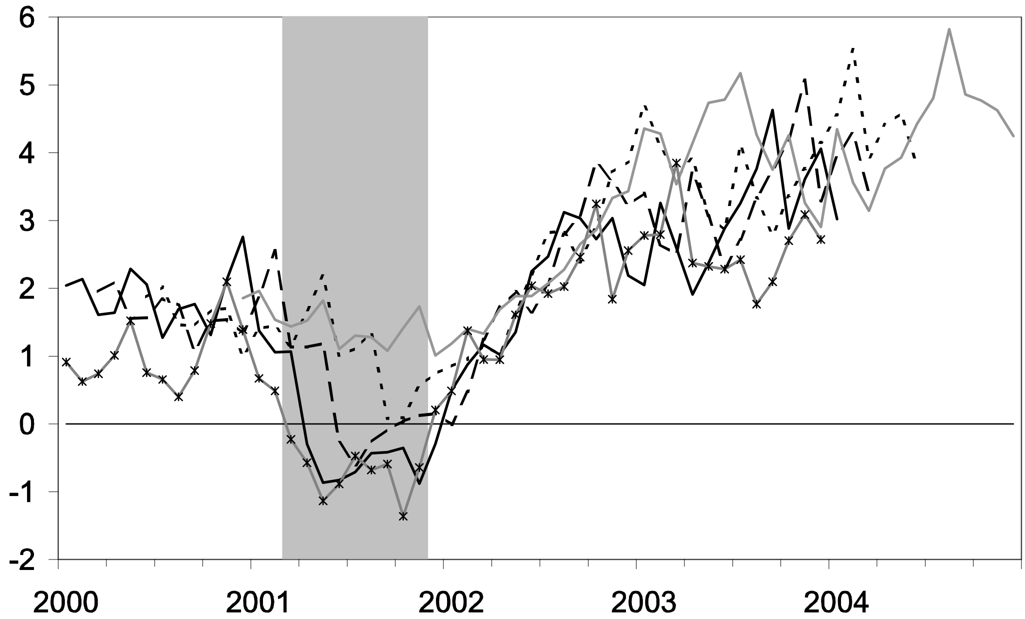

Figure 2: *Forecasts of Latent Variable.* Recursive out-of-sample forecasts from January 2000 to December 2003. The starred grey line represents the posterior mean of the latent variable. The straight black line is the 1 month-, the line with the long dashes the 3 months-, the line with the short dashes the 6 months-, and the straight grey line the 12-months-ahead forecast. The shaded area indicates the 2001 recession.

Table 1: Comparison of Qual VAR with Simple Probit

Forecast horizon	(posterior mean) RMSE		Classification errors	
	Qual VAR	Probit	Qual VAR	Probit
1 month	0.20	0.29	2	5
3 months	0.30	0.32	4	6
6 months	0.37	0.44	9	9
12 months	0.43	0.40	9	9

Note: Posterior mean of root mean squared error (RMSE) for QualVAR and RMSE for probit model for recursive out-of-sample forecasts from January 2000 to December 2003.

Table 2: Posterior Mean of the RMSE for Qual VAR and Simple VAR

Forecast horizon		Industrial production	Inflation	Interest rate	Term spread
1 month	Qual VAR	0.46	2.18	0.17	0.28
	Simple VAR	0.52	2.20	0.18	0.29
3 months	Qual VAR	0.56	2.14	0.52	0.63
	Simple VAR	0.63	2.17	0.57	0.66
6 months	Qual VAR	0.61	2.34	1.01	0.96
	Simple VAR	0.70	2.28	1.18	1.03
12 months	Qual VAR	0.62	2.46	1.79	1.37
	Simple VAR	0.65	2.37	2.16	1.52

Note: Recursive out-of-sample forecasts from January 2000 to December 2003.

Table 3: Posterior Mean of Difference in RMSE

Forecast horizon		Industrial production	Inflation	Interest rate	Term spread
1 month	RMSE diff.	12.09	0.85	3.69	2.14
	Pct. > 0	0.00	0.25	0.18	0.05
3 months	RMSE diff.	11.80	1.71	8.88	3.67
	Pct. > 0	0.00	0.22	0.04	0.05
6 months	RMSE diff.	12.73	-2.24	14.53	7.50
	Pct. > 0	0.00	0.83	0.00	0.00
12 months	RMSE diff.	4.89	-3.89	17.18	9.58
	Pct. > 0	0.00	0.90	0.00	0.00

Note: Posterior mean of difference in RMSE between the Qual VAR and the Simple VAR, and percentage of draws below zero for recursive out-of-sample forecasts from January 2000 to December 2003.