Forecasting Key US Macroeconomic Variables with a Factor-Augmented Qual VAR
Rangan Gupta
University of Pretoria
Eric Olson
West Virginia University
Mark E. Wohar
University of Nebraska at Omaha and Loughborough University
November 2015
In this paper, we first extract 8 factors from a monthly data set of 130 macroeconomic and financial variables. Then these extracted factors are used to construct a Factor-Augmented Qualitative VAR (FA-Qual VAR) model to forecast industrial production growth, inflation, the Federal funds rate and the term spread based on a pseudo real-time recursive forecasting exercise over an out-of-sample period of 1980:1-2014:12, using an in-sample period of 1960:1-1979:12. Short-, medium- and long-run horizons of one-, six, twelve- and twenty-four-month(s)-ahead are considered. The forecasts from the FA-Qual VAR is compared with that of a standard VAR model (comprising of output, prices, interest rate and the term spread), and that of a Qualitative VAR (Qual VAR) model (which includes the variables in the VAR and the latent business cycle index generated based on the information from the industrial production growth, inflation, the Federal Funds rate and the term spread). In general, we observe that the FA-Qual VAR tends to perform significantly better than the VAR and Qual VAR for the one-month-ahead and six-months-ahead forecast horizons for the key US variables under consideration. In other words, adding information from a large data set (through the use of factors) tend to produce forecasting gains at short- to medium-run horizons.

**JEL Classifications:** C32, C53, C55, E37, E47

**Keywords:** Vector Autoregressions, Business Cycle Turning Points, Factors, Forecasting
1. Introduction

In a seminal contribution, Sims (1980) proposed the Vector Autoregressive (VAR) model, which is essentially a system of equations, whereby in a specific equation, a specific variable is regressed on the past values of itself and past values other variable(s) in this system, with possible allowance for deterministic terms (e.g., constant and trend). By design, this linear framework treats all variables in the system as endogenous. Though atheoretical, the VAR model has been shown to perform exceptionally well when compared to various other econometric models, including its popular predecessor, namely the simultaneous equations approach of the Cowles Foundation (Dueker and Assenmacher-Wesche, 2010).

However, one criticism of the VAR forecasting is that macroeconomic variables tend not to behave as linear functions of their own past values around business cycle turning points. Given this issue, Hamilton (1989) developed the nonlinear Markov-switching model, in which coefficient switches are aligned with turning points of the business cycle. In these models, the regime switching mechanisms employed is usually based on a latent state variable or time-varying regime probabilities that have been specified as logistic functions of lagged endogenous variables. Furthermore, an additional form of nonlinear modelling of macroeconomic variables is by Leamer and Potter (2004), who, in turn, allow a latent business cycle index to alter the dynamics as the threshold variable in a threshold autoregressive model.

However, another alternative, as developed by Dueker (2005), is a new-type of regime switching VAR model, which is based on the idea that the regime is determined by an observed qualitative response variable that is modeled simultaneously within the VAR, referred to as a qualitative VAR (Qual VAR). Models of this form consider a qualitative
variable that is binary, and hence has two regimes such as the state of the business cycle. The advantage of this approach is that it allows one to understand which economic forces drive the regime switches. Dueker and Assenmacher-Wesche (2010) created a Qual VAR model which included a truncated normal latent business cycle index that is negative during NBER recessions and positive during expansions, over and above the standard variables of output, prices, interest rates and the term spread (difference between the 10-year government bond yield and the 3-month Treasury bill rate) comprising the VAR. These authors then apply the Qual VAR model to recursive out-of-sample forecasting of key US macroeconomic variables and find that the Qual VAR improves on out-of-sample forecasts from a standard VAR. Earlier Dueker (2005) had used a similar model with great success to predict ex ante the US recession of March, 2001. To the best of our knowledge, the papers by Dueker (2005) and Dueker and Assenmacher-Wesche (2010) are the only two papers till date that have used Qual VARs to carry out forecasting exercises. More recently, the Qual VAR approach has been used to analyse causality between interest rates and state of the business cycle (Nyberg, 2013), and the impact of unconventional monetary policy (Meinusch and Tillmann, forthcoming; Tillmann; 2015, forthcoming).

Understandably, the information content of the latent business cycle index generated from the Qual VAR would be contingent on the other variables comprising the Qual VAR. A presumption is that more variables in the Qual VAR are likely to bring in more information, and hence, a better description of the business cycle index, and in the process, produce superior forecasts of the key macroeconomic variables. Given this, we develop a factor-augmented Qual VAR (FA-Qual VAR), where we now incorporate (eight) factors extracted from a large monthly data set of 130 macroeconomic and financial variables, which excludes output, prices, interest rates and the term spread. We then use this FA-Qual VAR model to forecast output (growth of industrial production), prices (inflation based on the consumer
price index (CPI)), interest rates (the Federal Funds rate) and term spread of the US economy in a pseudo real-time recursive forecasting exercise over an out-of-sample period of 1980:1-2014:12, using an in-sample period of 1960:1-1979:12. The starting point of the out-of-sample period is in line with the wide-spread belief that the US economy underwent a major structural change in 1980 (Bekiros and Paccagnini, 2013). The forecasts from the FA-Qual VAR is compared with that of a standard VAR model comprising of output, prices, interest rate and the term spread, and that of a Qual VAR model which includes the variables in the VAR and the latent business cycle index generated based on the information from the industrial production growth, inflation, the Federal Funds rate and the term spread. Note that, to accommodate for the problem of overparametrization (often leading to poor forecasts) in the Qual VAR and the FA-Qual VAR when compared to the standard VAR, we use Bayesian shrinkage to ensure that all these models have the same in-sample fits and hence, a fair out-of-sample comparison, following the suggestions of Banbura et al., (2010). To the best of our knowledge, this is the first such attempt to develop and forecast with a FA-Qual VAR. The remainder of the paper is organized as follows: Section 2 presents a brief description of the Qual VAR model. Section 3 describes the data and presents the results from the forecasting exercise. Finally, section 4 concludes.

2. The Qualitative Vector Autoregressive (Qual VAR) Model

Suppose we observe a qualitative variable, $y \in \{0,1\}$, which is driven by a continuous latent variable, $y^*$, such that:

\[
y_t = 0 \text{ iff } y_t^* \leq 0 \\
= 1 \text{ iff } y_t^* > 0
\]  \hspace{1cm} (1)

along with:
\[
y_t^* = \Psi(L)y_{t-1} + \Gamma(L)X_{t-1} + \epsilon_t
\]
\[
\epsilon \sim N(0,1),
\]
where \(X_{t-1}\) is a set of explanatory variables (industrial production growth, inflation, the Federal Funds rate, the term spread in case of the Qual VAR or, for the FA-Qual VAR, the extracted factors and all the explanatory variables in the Qual VAR), 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. Then, as described in Dueker (2005) and Dueker and Assenmacher-Wesche (2010), a Qual VAR model with \(k\) variables and \(p\) lags is expressed as a standard VAR:

\[
\Phi(L)Y_t = \epsilon_t
\]

with \(\epsilon_t \sim \text{Normal}(0, \Sigma)\), and where \(Y_t\) is a \(k \times 1\) vector consisting of macroeconomic data (discussed in the above paragraph), \(X_t\), plus the latent business cycle turning point index \(y^*\); \(\Phi(L)\) is a set of \(k \times k\) matrices from \(L = 0, \ldots, p\), with the identity matrix at \(L = 0\), i.e., the VAR regression coefficients are in \(\Phi(L)\). The parameters that require conditional distributions for Markov Chain Monte Carlo (MCMC) estimation are \(\Phi, y^*\) and \(\Sigma\) (i.e., the covariance matrix), which in turn, involves a sequence of draws from the following conditional distributions, where superscripts indicate the iteration number:

VAR coefficients \(\sim\) Normal

\[
f \left( \Phi^{(i+1)} \mid \{y_t^{*(i)}\}_{t=1 \ldots T}, \{X_t\}_{t=1 \ldots T}, \Sigma^{(i)} \right)
\]

Covariance matrix \(\sim\) inverted Wishart

\[
f \left( \Sigma^{(i+1)} \mid \{y_t^{*(i)}\}_{t=1 \ldots T}, \{X_t\}_{t=1 \ldots T}, \Phi^{(i)} \right)
\]

Latent variable \(\sim\) truncated Normal

\[
f \left( y_t^{*(i+1)} \mid \Phi^{(i+1)}, \Sigma^{(i+1)}, \{y_{t}^{*(i)}\}_{t < t}, \{y_{k}^{*(i)}\}_{k > t}, \{X_t\}_{t=1 \ldots T} \right)
\]
Conditional on a set of values for $y^*$, $\Phi$ are normally distributed. While, $\Sigma$ is part of a normal-inverted Wishart conjugate pair with $\Phi$. Finally, each observation of $y^*$, has a truncated normal distribution, where it is not allowed to be negative (positive) during expansions (recessions). Further details on these conditional distributions and prior distributions can be found in the Appendix of Dueker and Assenmacher-Wesche (2010).

3. Data and Forecasting Results

Besides the data on industrial production, CPI, the Federal Funds rate and the term spread, our monthly data set over the period of 1960:1-2014:12 includes 130 other macroeconomic and financial variables. The data set (FRED-MD) developed by McCracken and Ng (2015) is available for download from the following link on the website of the Federal Reserve Bank of St. Louis: http://research.stlouisfed.org/econ/mccracken/sel. All of the 130 variables are transformed to induce stationarity before extracting 8 factors, as determined by the tests of Bai and Ng (2003). The reader is referred to the Appendix of McCracken and Ng (2015) for complete details about the data transformation. Note that the MCMC approach of the Qual VAR requires data to be stationary of the other explanatory variables, hence, we also use growth rate of industrial production and month-on-month CPI-based inflation rate. The extracted factors, used in the FA-Qual VAR, are stationary by design. Data transformations imply that our effective sample starts from 1960:2-2014:12, with the end date ensuring that our data set is a balanced one, and the starting point of being driven purely by data availability. Note that, the qualitative variable is a binary 0/1 variable that denotes recessions and expansions, with switches at business cycle turning points, with the recession and expansion classification coming from the NBER.

As discussed in the introduction, our in-sample ends in 1979:12, with the out-of-sample
period starting in 1980:1-2014:12, over which the models are recursively estimated to produce forecasts at horizons \( (h) \) of one-, six-, twelve-, and twenty four-month(s)-ahead. At this stage, it is important to clearly lay out the structure of our pseudo out-of-sample forecasting exercise for the VAR, Qual VAR and the FA-Qual VAR. For the FA-Qual VAR, we first determine the choice of eight factors based on the in-sample (1960:2-2014:12) by leaving out the variables in the VAR (i.e., industrial production growth, inflation, Federal Funds rate and the term spread). Keeping the number of factors fixed at 8, we recursively estimate the factors adding one observation at a time over the out-of-sample period (1980:1-2014:12). Then these recursively estimated factors are fed into the VAR model, along with the recession/expansion dummy, to extract the corresponding underlying latent business cycle indicator also in a recursive way by adding one observation at a time for the variables of interest and the qualitative variable. For the Qual VAR, we follow the same approach, but now, we only need to recursively add the information for the variables of interest and the qualitative variable. For the FA-Qual VAR and Qual-VAR, following Dueker and Assenmacher-Wesche (2010), for each estimation, we use 1200 iterations from which the first 300 are discarded to allow for convergence towards the posterior distribution. Figure A1 in the Appendix of this paper plots the two (standardized to have unit variance) latent business cycle indicator generated under the Qual VAR and the FA-Qual VAR. As can be seen, the latent business cycle indicators behave quite similarly, but the latent indicator obtained from the Qual VAR tends to predict stronger evidence of expansions for many periods, while that from the FA-Qual VAR predicts relatively deeper recessions. Finally, the standard VAR, the Qual VAR and the FA-Qual VAR are now estimated recursively over the out-of-sample period to produce the forecasts at \( h = 1, 6, 12 \) and 24 over the out-of-sample period. All the models are estimated based on a lag-length \( (p) \) of 3, with it being determined based on the Akaike
Information Criterion (AIC) for the standard VAR’s estimation over the in-sample period. The lag-length is kept fixed over the out-of-sample period while generating forecasts. Note that we can only conduct a pseudo real-time forecasting analysis when a business cycle indicator is involved, since there is considerable time lag with which the NBER releases information that a turning point has occurred. Our data-vintage corresponds to September, 2015.

An additional issue that needs to be discussed now, is the issue of overparameterization in the Qual VAR and the FA-Qual VAR relative to the standard VAR. To address this concern, we impose the standard Minnesota prior, as developed by Litterman (1986), on the parameters of the Qual VAR and the FA-Qual VAR. The Minnesota prior imposes restrictions on the coefficients of longer lags by assuming that these are more likely to be near zero than the coefficient on shorter lags. However, if there are strong effects from less important variables, the data can override this assumption. The restrictions are imposed by specifying normal prior distributions with zero means and small standard deviations for all coefficients with the standard deviation decreasing as the lags increases. The exception to this, however, is the coefficient on the first own lag of a variable, which has a mean of unity. But, if the variables in the VAR are stationary (which happens to be in our case), one needs to also impose a zero prior mean on the coefficient of the first own lag of a variable. The reader is referred to Litterman (1986) and Banbura et al., (2010) for the technical details, which is quite well-known in the forecasting literature now, and hence has been omitted from the discussion here. In the Minnesota prior-settings, a hyperparameter is used to determine how prior beliefs relate to the information contained in the data. More precisely, this hyperparameter controls the overall tightness of the prior distribution around the prior mean. Banbura et al., (2010), argue that this tightness-hyperparameter should reflect the size of the system, i.e., as the number of variables increases, the parameters
should shrink to avoid overfitting. This, in turn, is achieved by setting this hyperparameter to match the in-sample fit as the benchmark VAR. Understandably, the VAR has an infinite value for this hyperparameter with the corresponding values for the Qual VAR and the FA-Qual VAR being 0.2459 and 0.0877 respectively, to ensure an in-sample fit of 0.3584 – as obtained for the VAR in our case.

Table 1 presents the forecasting results for industrial production growth, inflation, the Federal funds rate and the term spread at $h=1, 6, 12$ and 24 emanating from the VAR, Qual VAR and the FA-Qual VAR. The entries corresponding to the FA-Qual VAR for each variable are MSFE of the FA-Qual VAR relative to the Random Walk (RW) model. Hence a value less than one is indicative of the superior forecasting performance of the FA-Qual VAR in comparison to the RW model, which in turn, is found to be the case for all the four variables at all horizons, except for the case of industrial production at $h=12$. Entries corresponding to VAR and Qual VAR for each variable are MSFE of the VAR and Qual VAR relative to the FA-Qual VAR model. Thus, in cases where these values are greater than one, it indicates that the FA-Qual VAR outperforms the VAR and Qual VAR. As can be seen, this is consistently the case, barring some exceptions, at short- to medium-run horizons, i.e., $h=1$ and 6 respectively. For the case of the industrial production growth, the FA-Qual VAR outperforms the VAR even at $h=6$ and 12 as well, while for the Federal Funds rate the VAR does slightly better than the FA-Qual VAR at the one-month-ahead horizon. Besides these cases, the VAR and Qual VAR consistently outperforms the FA-Qual VAR at long- to very long-run horizons, i.e., $h=12$ and 24. It is however, important to point out that for $h=12$ and 24, barring the case of the term spread at the twenty-four-months-ahead horizon, the Qual VAR outperforms the VAR. So taking the performance of the Qual VAR and the FA-Qual VAR together, our results highlight the importance of modeling nonlinearity through the use of the latent business cycle indicator.
Note that the FA-Qual VAR nests all the other models, i.e., the RW, VAR and Qual VAR. Given this, we use McCracken’s (2007) powerful ENC-NEW test statistic to determine whether the scenarios in which the FA-Qual VAR outperforms the other models is significant or not.\(^1\) As indicated by the bold entries in the table, the ENC-NEW statistic is significant at the 5 percent level for all the 32 cases. So in sum, the FA-Qual VAR tends to perform better than the VAR and Qual VAR at short- to medium-run forecasting horizons.

[Insert Table 1 Here]

4. Conclusions

Though atheoretical, the linear vector autoregressive (VAR) model has been shown to perform exceptionally well when compared to various other econometric models. However, one criticism of the VAR forecasting is that macroeconomic variables tend not to behave as linear functions of their own past values around business cycle turning points. Given this, Dueker (2005) developed a new-type of regime switching VAR model, called the qualitative VAR (Qual VAR), which, in turn, is based on the idea that the regime is determined by an observed qualitative response variable that is modeled simultaneously within the VAR. Recently, Dueker and Assenmacher-Wesche (2010) created a Qual VAR model which included a truncated normal latent business cycle index that is negative during NBER recessions and positive during expansions, over and above the standard variables of output, prices, interest rates and the term spread (difference between the 10-year government bond yield and the 3-month Treasury bill rate) comprising the VAR. These authors then apply the Qual VAR model to recursive out-of-sample forecasting of key US macroeconomic variables and find that the Qual VAR improves on out-of-sample forecasts from a standard VAR.

\(^1\) The ENC-NEW statistic tests whether the restricted forecast results from the VAR or the Qual VAR encompass the unrestricted forecast results from the FA-Qual VAR. If so, then the additional variables in the FA-Qual VAR provide no further predictive power than the VAR and the Qual VAR models. A rejection of the null hypothesis means that the unrestricted model forecasts are not encompassed by the restricted results.
Realizing that, the information content of the latent business cycle index generated from the Qual VAR would be contingent on the other variables comprising the Qual VAR, we develop a factor-augmented Qual VAR (FA-Qual VAR), where we now incorporate (eight) factors extracted from a large monthly data set of 130 macroeconomic and financial variables, which excludes output, prices, interest rates and the term spread. The presumption is that more variables in the Qual VAR are likely to bring in more information, and hence, a better description of the business cycle index, and in the process, produce superior forecasts of the key macroeconomic variables. Then we use this FA-Qual VAR model to forecast industrial production growth, inflation, the Federal Funds rate and the term spread of the US economy in a pseudo real-time recursive forecasting exercise over an out-of-sample period of 1980:1-2014:12, using an in-sample period of 1960:1-1979:12. The forecasts at horizons of one-, six-, twelve- and twenty-four-month(s)-ahead from the FA-Qual VAR is compared with that of a standard VAR model comprising of output, prices, interest rate and the term spread, and that of a Qual VAR model which includes the variables in the VAR and the latent business cycle index generated based on the information from the industrial production growth, inflation, the Federal Funds rate and the term spread. In general, we observe that the FA-Qual VAR tends to perform significantly better than the VAR and Qual VAR for the one-month-ahead and six-months-ahead forecast horizons for the key US variables under consideration. In other words, adding information from a large data set (through the use of factors) tend to produce forecasting gains at short- to medium-run horizons. In addition, for the longer horizons, the Qual VAR generally tends to outperform the VAR. Hence, when we consider the various horizons, the importance of modeling nonlinearity through the use of a latent business cycle indicator, either based on a Qual VAR or a FA-Qual VAR cannot be denied. As part of future research, it would be interesting to compare the (FA-)Qual VAR with other
nonlinear frameworks like the Markov-switching and threshold VARs. In addition, given that the Qual VARs are estimated using Bayesian methods, we can also analyse density forecasts.
References
**Table 1.** Forecasting results from VAR, Qual VAR and FA-Qual VAR: 1980:1-2014:12

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model</th>
<th>$h$</th>
<th>1</th>
<th>6</th>
<th>12</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial Production Growth</td>
<td>FA-Qual VAR</td>
<td>0.5975</td>
<td>0.6727</td>
<td>1.0300</td>
<td>0.9353</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VAR</td>
<td>1.1409</td>
<td>1.0456</td>
<td>1.0314</td>
<td>1.0262</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qual VAR</td>
<td>1.0117</td>
<td>1.0454</td>
<td>0.9538</td>
<td>0.9597</td>
<td></td>
</tr>
<tr>
<td>Inflation</td>
<td>FA-Qual VAR</td>
<td>0.6099</td>
<td>0.7187</td>
<td>0.9002</td>
<td>0.8359</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VAR</td>
<td>1.0375</td>
<td>1.0566</td>
<td>0.9414</td>
<td>0.9657</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qual VAR</td>
<td>1.0276</td>
<td>1.0354</td>
<td>0.9048</td>
<td>0.9307</td>
<td></td>
</tr>
<tr>
<td>Federal Funds Rate</td>
<td>FA-Qual VAR</td>
<td>0.5059</td>
<td>0.7262</td>
<td>0.7958</td>
<td>0.7246</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VAR</td>
<td>0.9910</td>
<td>1.1872</td>
<td>0.9289</td>
<td>0.9445</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qual VAR</td>
<td>1.0484</td>
<td>1.1136</td>
<td>0.9177</td>
<td>0.9366</td>
<td></td>
</tr>
<tr>
<td>Term Spread</td>
<td>FA-Qual VAR</td>
<td>0.4593</td>
<td>0.7733</td>
<td>0.6586</td>
<td>0.6297</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VAR</td>
<td>1.0230</td>
<td>1.2446</td>
<td>0.9782</td>
<td>0.9562</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qual VAR</td>
<td>1.0547</td>
<td>1.1456</td>
<td>0.9596</td>
<td>0.9588</td>
<td></td>
</tr>
</tbody>
</table>

Note: Entries corresponding to FA-Qual VAR for each variable are MSFE of the FA-Qual VAR relative to the Random Walk (RW) model; Entries corresponding to VAR and Qual VAR for each variable are MSFE of the VAR and Qual VAR relative to the FA-Qual VAR model. Entries in bold indicate significance of McCracken’s ENC-NEW statistic at the 5 percent level.
APPENDIX:

Figure A1. Plot of the Latent Business Cycle indicator for the Qual VAR and the FA-Qual VAR