Economics Bulletin

Volume 34, Issue 3

Forecasting Exchange Rates using Bayesian Threshold Vector Autoregressions

Florian Huber
Vienna University of Economics and BA

Abstract

In this paper we assess the predictive abilities of a Bayesian threshold vector autoregression (B-TVAR) to forecast the EUR/USD exchange rate. By introducing stochastic search variable selection priors (SSVS), we account for the inherent model uncertainty when it comes to modeling exchange rates. Our results suggest that, by applying Bayesian methods to the TVAR, it is possible to improve upon the random walk forecast. Surprisingly, we even managed to outperform the naive benchmark model in short-term forecasting, where the gains in terms of predictive ability are substantial.

The author would like to thank Jesus Crespo-Cuaresma, Philipp Piribauer, Octavio Fernandez-Amador and Gianni Amisano for helpful comments.
Citation: Florian Huber, (2014) "Forecasting Exchange Rates using Bayesian Threshold Vector Autoregressions", *Economics Bulletin*, Vol. 34 No. 3 pp. 1687-1695.
Contact: Florian Huber - fhuber@wu.ac.at.
Submitted: June 09, 2014. **Published:** August 06, 2014.

1 Introduction

Forecasting exchange rates has been a nightmare for economists since the beginning of the eighties, when the influential paper by Meese and Rogoff (1983) showed that no structural model was able to beat the simple random walk in terms of predictive accuracy. In the following ten years, most studies dealing with exchange rate prediction concluded that it is not possible to improve upon the no-change forecast. The success of the random walk was challenged in the beginning of the nineties, when MacDonald and Taylor (1993) were able to outperform the random walk by allowing for cointegration between the nominal exchange rate and macroeconomic fundamentals. A few years later, Mark (1995) showed that the unpredictability of exchange rates is mainly a short-term phenomenon, implying that for time horizons greater than one year, the models considered outperformed the random walk benchmark by a large margin.

Comparatively few attempts have been made to apply non-linear models to exchange rate forecasting. In one of the rare contributions, Clements and Smith (1997) investigated the forecasting performance of self-exciting threshold autoregressive (SETAR) models for modeling exchange rates. They concluded that non-linear models tend to perform well when evaluated conditional on the state of nature, but not when the full sample is taken under consideration. As a generalization of the univariate threshold autoregressive model, threshold vector autoregressions (TVARs) were introduced by Tsay (1998). Conditional on the regime, TVARs possess the same features and shortcomings as standard VARs. Consequently, they also suffer from the well known curse of dimensionality, which translates into severe overparameterization and weak out-of-sample forecasting performance.

Our contribution to the literature is twofold. First, we extend the Markov-Chain Monte Carlo (MCMC) algorithm outlined in So and Chen (2003) to the multivariate threshold case. Second, we solve the curse of dimensionality through the introduction of Bayesian shrinkage priors. More specifically, we utilize the stochastic search variable selection prior (SSVS) put forward by George and McCulloch (1993) and the subsequent developments for VARs by George et al. (2008) to increase the predictive performance of the TVAR. Our results suggest that it is possible to improve upon simpler, frequentist models considered in earlier studies. More importantly, the Bayesian TVAR outperforms the random walk on all time horizons considered. However, the largest improvements can be found for time horizons greater than one month.

2 Data and Methodology

2.1 Data Overview

Our dataset ranges from 1990M07 to 2012M07. The money supply for the home country, which is the Euro Area (EA), is the logarithm of the M2 monetary base for the EA and the logarithm of the US M2 monetary base for the United States. The short term interest rate is the 3-month interbank rate for the US and the EA short term interest rate is approximated using the German 3-month interbank rate. The logarithm of the industrial output for the US and Germany is used as proxy for GDP. The expected future inflation differential is approximated by using the difference between 10-year US treasury yields and 10-year German Bund yields. Table 1 provides an overview of the dataset employed with the corresponding Datastream codes.

Table 1: Data Overview

Description	Datastream ID	Mean	St. Dev.	Min	Max
EUR/USD* exchange rate	USEURSP	0.194	0.143	−0.171	0.455
M2 money supply* (EA)	EMM1....B	7.689	0.519	6.873	8.508
M2 money supply* (US)	USM1....B	7.120	0.230	6.682	7.749
3-month interbank rate (EA)	EMINTER3	1.289	0.739	−0.699	2.470
3-month interbank rate (US)	USINTER3	1.031	0.986	−1.427	2.148
Industrial Production* (Germany)	BDIPTOT.G	4.554	0.095	4.386	4.750
Industrial Production* (US)	USIPTOT.G	4.416	0.158	4.099	4.613
10y government bond yields (EA)	EMGBOND.	1.697	0.364	1.149	2.410
10y government bond yields (US)	USGBOND.	1.706	0.266	0.798	2.185

Note: Asterisks indicate that the variables are entering the model in logarithms. Final model specification takes all the variables as log country-deviations.

2.2 Threshold VAR

Following Tsay (1998), the TVAR with R regimes is given by

$$y_t = \sum_{r=1}^{R} \left[(\psi_{r,0} + \psi_{r,1} t + \sum_{i=1}^{p} \Phi_{r,i} y_{t-i} + u_{r,t}) \times I(z_{t-k} \in (\gamma_{r-1}, \gamma_r)) \right] \tag{1}$$

where

$$I(z_{t-k}) = 1 \Leftrightarrow z_{t-k} \in (\gamma_{r-1}, \gamma_r) \tag{2}$$

y_t is a $K \times 1$ vector of explanatory variables, $u_{r,t} \sim \mathcal{N}(0, \Sigma_r)$ denotes the vector of error terms, Σ_r is the variance-covariance matrix and $\Phi_{r,i}$ is the $K \times K$ matrix of coefficients in regime $r = 1, ..., R$. $\psi_{r,0}$ and $\psi_{r,1}$ denote the coefficients on the constant and trend in each regime, respectively. p determines the number of lags of the endogenous variables. Note that $z \subseteq Z$ is the threshold variable, where Z is a matrix of possible threshold variables and γ_r defines the threshold parameter. In the following discussion we restrict Z to be a subset of y. Finally, $k = 1, .., K$ determines the lags of the threshold variables. For the subsequent discussion, it proves to be convenient to work with the regime specific coefficient vector $\Pi_r := (\psi'_{r,0}, \psi'_{r,1}, \text{vec}(\Phi_{r,1})', ..., \text{vec}(\Phi_{r,p})')$.

2.3 Priors and Estimation Strategy

Priors on Π_r and Σ_r

The aforementioned problem of overfitting in TVARs is reduced by means of the SSVS prior, which selects the appropriate degree of shrinkage for the parameters in equation (1). Conditional on the regime, we assume a mixture normal prior on each VAR coefficient:

$$\Pi_{r,j}|\delta_{r,j} \sim (1 - \delta_{r,j})\mathcal{N}(0, \omega^2_{r,0j}) + \delta_{r,j}\mathcal{N}(0, \omega^2_{r,1j}) \tag{3}$$

where $\delta_{r,j}$ is a dummy random variable which corresponds to coefficient j in regime r. It equals one if the coefficient is to be included in the model and zero if it is omitted. This implies drawing from one of the normal priors in equation (3), where $\omega^2_{r,0j}$ is the prior variance on the coefficient for the first normal, which is set to a value close to zero, implying an informative prior and effectively shrinking the coefficient towards zero. The prior variance $\omega^2_{r,1j}$ for the second normal prior is set to a large value, leading to an uninformative prior on coefficient j. The exact specification of the hyperparameters is done in a default semi-automatic fashion, proposed by George et al. (2008), where the $\omega^2_{r,0j}$ and $\omega^2_{r,1j}$ are scaled using the OLS standard deviation of the parameter in question.

The intuition behind the SSVS prior is that, depending on whether $\delta_{r,j}$ equals zero or one, a variable is excluded or included in the model. This is done by imposing a dogmatic prior on $\Pi_{r,j}$ if $\delta_{r,j} = 0$ and a diffuse prior if $\delta_{r,j} = 1$. The first case implies that the corresponding coefficient could be safely regarded as zero, implying that the variable is not included in the model. Averaging the draws of the $\delta_{r,j}$ leads to posterior inclusion probabilities for variable j in regime r.

For the elements of Σ_r, we impose the following prior structure

$$[\Sigma_r]_{i,j} \sim \begin{cases} \mathcal{G}(a_1, a_2) & \text{if } i = j \\ \mathcal{N}(0, \underline{V}) & \text{if } i \neq j \end{cases} \tag{4}$$

where the hyperparameters a_1 and a_2 are set to small values such that the prior is rendered noninfluental. We specify the prior variance on the off-diagonal elements as $\underline{V} = \text{diag}(5^2)$, which again corresponds to a non-informative prior choice.

Priors on γ

For the threshold parameter, we follow the literature on Bayesian threshold autoregressions (Geweke and Terui, 1993; So and Chen, 2003) and use a Normal prior on γ_r, formally:

$$\gamma_r \sim \mathcal{N}(\underline{\mu}_z, \underline{\sigma}_z) \tag{5}$$

where $\underline{\mu}_z$ and $\underline{\sigma}_z$ denote the prior mean and variance on the threshold, respectively. In the following application we set the prior mean equal to zero and the variance equal to 10, implying a fairly diffuse prior on the threshold.

MCMC Algorithm

Conditional on the regime, estimation of the model in Equation 1 can be done in a straightforward fashion by employing the MCMC algorithm outlined in George et al. (2008). In the following discussion, we restrict the number of regimes in our analysis to two. Estimation of γ_r is done by including a random walk Metropolis step in our MCMC algorithm. The law of motion for γ_r is given by

$$\gamma_r^* = \gamma_r^{(a)} + \sigma_s e \tag{6}$$

where γ_r^* is the proposed value for γ_r, $\gamma_r^{(a)}$ denotes the last accepted draw and $e \sim \mathcal{N}(0,1)$. σ_s is a scaling factor set such that the acceptance rate defined below is between 20 and 40 percent. The draw γ_r^* is then evaluated by by comparing the conditional posterior of $p(\gamma_r^*|y, \Pi_r, \Sigma_r, \delta_r, k, z)$ with the posterior at the last accepted draw for $\gamma_r^{(a)}$. The probability of accepting a draw is

$$\alpha(\gamma_r^*|\gamma_r^{(a)}) = \min\left(\frac{\prod_{r=1}^2 p(y|\gamma_r^*, \Pi_r, \Sigma_r, \delta_r, k, z)p(\gamma_r^*)}{\prod_{r=1}^2 p(y|\gamma_r^{(a)}, \Pi_r, \Sigma_r, \delta_r, k, z)p(\gamma_r^{(a)})}, 1\right) \tag{7}$$

One consequence of this is that γ_r's with higher posterior density than the old parameters will be accepted in any case, whereas draws with lower posterior density will be accepted with a probability proportional to the ratio of the posterior at γ_r^* and $\gamma_r^{(a)}$.

For the lag of the threshold variable k, Chen and Lee (1995) show that $p(k|y, \Pi_r, \Sigma_r, \delta_r, \gamma_r)$

follows a multinomial distribution with a probability vector

$$p(k = d|y, \Pi, \Sigma, \delta, \gamma_r) = \frac{\prod_{r=1}^{2} p(y|\gamma_r, \Pi_r, \Sigma_r, \delta_r, k = d, z)}{\sum_{d=1}^{d_{max}} \prod_{r=1}^{2} p(y|\gamma_r, \Pi_r, \Sigma_r, \delta_r, k = d, z)} \quad (8)$$

This part is easily implemented as an additional step in the Gibbs sampler.

The last ingredient is the specific threshold variable chosen. To account for the uncertainty regarding the threshold variable, we estimate the model for all $z \in Z$ and choose the one which yields the highest marginal likelihood approximated by the Bayesian Information Criterion (BIC).

The model in (1) is estimated in country differences, implying relative deviations of the corresponding covariates in the system. Furthermore we included four lags for the explanatory variables in the following analysis, permitting the SSVS prior to select the most appropriate lag length and shrinking the coefficients on the less important lags towards zero.

3 Results

We rely on the following forecasting design: As an initial estimation period, 1990M01 to 2010M07 (246 Observations), is used. Then we perform rolling regressions, where the estimation window is expanded for one period each step, keeping the starting point (1990M01) fixed. The forecasting period is from 2010M07 to 2012M07 (24 observations).

Our goal is to show that the introduction of Bayesian methods and non-linearities in the modelling framework tends to exhibit a positive effect on the forecasting accuracy. Consequently, we benchmark the Bayesian TVAR (B-TVAR) against its linear counterparts and the random walk. For the linear counterparts, we have decided to include the SSVS-BVAR, a simple BVAR with a Minnesota prior and some simple, univariate models like the autoregressive model and the random walk with drift. In addition, we also show the results using a frequentist TVAR.

The mean of the predictive density is used as a point estimator for the forecasts. Comparison is then done using the well known root mean square error (RMSE). Judging the predictive accuracy by RMSE implies that uncertainty about the point forecasts is omitted. Thus we also rely on the log predictive likelihood (PL) as a Bayesian standard measure to compare density forecasts. The log predictive likelihood is defined as

$$PL := \sum_{t=t_0}^{T-h} \log \left[p(e_{t+h} = \tilde{e}_{t+h}|D_t) \right] \quad (9)$$

where $\tilde{e}_{t+h}$ is the actual realization of the exchange rate and D_t denotes the available information until time t. Higher PLs imply stronger predictive capabilities.

Table 2 presents the results of our forecasting exercise. The values presented are the RMSEs of the corresponding models relative to the RMSE of the naive benchmark model. Values in parentheses represent the log-predictive likelihood for each model. M denotes the BVAR with the Minnesota prior, SSVS is the BVAR with the SSVS prior, B-TVAR denotes the B-TVAR with the SSVS prior specification and TVAR is the threshold VAR estimated by maximum likelihood. AR and RWD denote the autoregressive model and the random walk with drift respectively. The last column shows the absolute RMSE of the random walk forecast.

Table 2: Relative Performance $RMSE_i/RMSE_{RW}$

	B-TVAR	SSVS	M	TVAR	AR	RWD	RW
1 M	**0.8503**	0.9540	0.9050	0.9824	1.0094	1.0019	0.0357
	(**48.9120**)	(46.7572)	(48.1426)	-	-	-	-
3 M	**0.8129**	0.8724	0.8929	0.9552	0.9482	1.0063	0.0486
	(**40.5917**)	(39.6345)	(39.0596)	-	-	-	-
6 M	0.9323	**0.8406**	0.8641	0.8801	0.9688	1.0070	0.0498
	(38.3456)	(41.6703)	(**41.9071**)	-	-	-	-
12 M	**0.7720**	0.8788	0.8652	0.8713	0.8988	1.0190	0.0815
	(**32.1721**)	(22.8304)	(22.8193)	-	-	-	-

Note: The figures refer to the relative RMSE of a model to the random walk RMSE. Log predictive likelihoods in parentheses. Results based on rolling forecasts over the time period 2010M01-2012M07. M stands for the Minnesota prior BVAR, SSVS stands for the SSVS prior BVAR, B-TVAR stands for the threshold vector autoregressive model with SSVS prior specification and TVAR is a standard threshold VAR estimated using maximum likelihood. AR and RWD refer to the autoregressive model of order one and random walk with drift respectively. RW is the absolute RMSE of the random walk forecast. Bold figures refer to the lowest value across models for the log EUR/USD exchange rate.

As can be seen in Table 2, all of the multivariate models considered perform quite well in the forecasting exercise, beating the random walk on every time horizon considered. Except for the six step ahead forecasts, the B-TVAR managed to outperform its peers by a large margin. Especially on the three and twelve month time frames, the performance is exceptionally strong, improving upon the random walk benchmark by approximately 15 to 23 percent. Comparison of the non-linear models with their linear competitors points to the fact that non-linearities tend to increase the predictive capabilities of the models considered, especially on longer time frames. Note that the results above suggest that even at the critical one-month-ahead horizon, the B-TVAR exhibits exceptional predictive capabilities, outperforming the random walk by nearly 15 percent. The B-TVAR also outperforms its frequentist counterpart on most time horizons considered. Especially on the three- and twelve month time horizons the improvement in terms of RMSE is quite large, revealing that the SSVS-prior seems to work well in that environment.

Comparison of the log predictive scores suggests that the models which excel when judged by means of their point forecasts are also the ones that succeed in the density forecasting exercise. More specifically, the B-TVAR outperforms its peers on the one, three- and twelve month time horizons in terms of log predictive scores, which is consistent with the findings above.

4 Conclusions

This study has shown that the introduction of model uncertainty and non-linearity in the context of exchange rate forecasting helps to improve upon simpler models considered in the traditional exchange rate forecasting literature. The SSVS prior allows to account for model uncertainty in a flexible way. One implication of this prior structure is that it serves as an automatic model selection device, applying shrinkage on the coefficients where it is needed. As a consequence of the hierarchical model structure, the required input from the researcher is minimized.

Our results indicate that it is possible to improve upon the random walk benchmark, even at the critical one-step-ahead horizon. The B-TVAR with SSVS prior beats all other models for most time horizons considered, both in terms of the point forecasts and the predictive densities. Especially the strong performance at the short-time horizons points towards the fact that it pays of to introduce the notions of model uncertainty and non-linearity in the modelling framework.

References

Chen CW, Lee JC. 1995. Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* **16**: 483–492.

Clements MP, Smith J. 1997. A Monte Carlo study of the forecasting performance of empirical SETAR models .

George EI, McCulloch RE. 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**: 881–889.

George EI, Sun D, Ni S. 2008. Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* **142**: 553–580.

Geweke J, Terui N. 1993. Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis* **14**: 441–454.

MacDonald R, Taylor MP. 1993. The monetary approach to the exchange rate: Rational expectations, long-run equilibrium, and forecasting. *IMF Staff Papers* **40**: 89–107.

Mark NC. 1995. Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review* **85**: 201–18.

Meese RA, Rogoff K. 1983. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics* **14**: 3 – 24. ISSN 0022-1996.

So MK, Chen CW. 2003. Subset threshold autoregression. *Journal of Forecasting* **22**: 49–66.

Tsay RS. 1998. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**.