Exchange Rate Predictability and Dynamic Bayesian Learning

Joscha Beckmann[1] Gary Koop[2] Dimitris Korobilis[3]
Rainer Schüssler[4]

[1]University of Greifswald [2]University of Strathclyde [3]University of Essex
[4]University of Rostock

Background

- Begin with overview of the research area I work in and why it might be of interest for financial analysts and macroeconomists
- Bayesian econometrics has enjoyed huge increase in popularity because of its ability to deal with two issues:
- Model uncertainty/model switching
- Big Data
- To explain these in a general sense, let me begin with a regression

Statistical Challenges in Modern Macroeconomics and Finance

- Many questions of interest to financial/macro economist involve regression
- y_t – dependent variable(s)
- x_t – explanatory variables
- Regression:

$$y_t = x_t'\beta + \varepsilon_t$$

- Challenge 1: x_t might contain lots of explanatory variables (most of which are probably unimportant)
- Challenge 2: y_t might contain lots of dependent variables
- E.g. portfolio analysis might involve stock returns on 500 companies in S&P500

Why are these things "challenges"?

- Models often have large number of parameters relative to number of observations
- E.g. large macroeconomic Vector Autoregressive (VAR) models: y_t is $n \times 1$, $n > 100$
- VARs have n equations and each has p lags of all the variables
- 100 equations, each with $p = 12$ (monthly) lags, leads to 1200 right hand side variables in each equation
- Over-parameterization and over-fitting.
- With so many dimensions to fit, somewhere you will fit the noise in data rather than pattern
- Consequence: Apparently good sample fit, but poor forecasting
- Computation: Bayesian analysis (MCMC) can be very slow

Dealing with Big Data

- Various Bayesian approaches:
- Prior shrinkage (hierarchical priors: LASSO, Horseshoe, SSVS, etc.) — machine learning methods
- Compress the data and then work with smaller model (principal components/factor models, random compression)
- Instead of working with large Big Data model, work with many smaller models and then average/select between them
- It is this last approach I will use today
- This leads to the issue of model uncertainty

Model Uncertainty

- What to do when you have many models?
- Conventional approaches: Bayesian Model Averaging (BMA) or Bayesian Model Selection (BMS)
- But in unstable times (financial crisis/eurozone crisis) the idea that there is one best model for all times is unrealistic
- Recent developments in Dynamic Model Averaging (DMA) or Dynamic Model Selection (DMS)
- E.g. best model to use to construct a financial portfolio may change over time (model switching).
- These dynamic model switching concepts are embedded in the remainder of this talk and are key to obtaining good empirical results

Exchange Rate Modelling

- Forecasting exchange rates is difficult
- Meese-Rogoff puzzle: hard to beat a random walk
- Many econometric approaches have been tried.
- These differ in following ways:

1. Which predictors are used
2. VARs (cross-section of ex rates for many countries) vs. single equation (e.g. regression or AR)
3. Treatment of model uncertainty (e.g single model or many; dynamic model selection vs. static)
4. Whether parameters (coefficients and volatilities) are constant over time or not

What We Do (Statistics)

- Develop statistical approach that allows for a general treatment of each of these categories.
- Use monthly FX data for G10 countries of most traded currencies along with four exogenous predictors
- Most flexible models is a 9 dimensional TVP-VAR with stochastic volatility involving exogenous predictors
- Within this hundreds of restricted models.
- Dynamic model selection: make specification choices in data-based fashion
- Algorithm automatically makes choices about predictors, VAR properties, rate of model switching or parameter change
- Does so in dynamic manner, allowing for different forecasting models to be used at different points of time.

What We Do (Empirical)

- Forecast 9 exchange rates (G10 all relative to US dollar): AUD, CAD, EUR, JPY, NZD, NOK, SWK, SWF and GBP
- Long data set: 1973M1 through 2016M12 involving this cross-section of exchange rates
- Short data set: 1986M1 through 2016M12 also includes UIP, STOCKS, INT DIFF, OIL
- Robustness check using more predictors (not all real-time)
- Evaluate forecasting performance of our approach in two ways:
- Viewpoint of statistician (predictive log likelihoods)
- Viewpoint of investor building an FX portfolio (economic utility)
- Relative to random walk benchmark:
- Find small gains from statistician's point of view
- Appreciable gains from investor's point of view

Large TVP-VARs

- y_t is vector containing observations on M time series variables
- TVP-VAR is:

$$y_t = Z_t \beta_t + \varepsilon_t$$

- Z_t defined to contain intercept lags of all the dependent variables
- Note Z_t is $M \times k$ where $k = M(1 + pM)$
- VAR coefficients evolve according to:

$$\beta_{t+1} = \beta_t + u_t$$

- Main results use $M = 9$, $p = 6$, then $k = 495$
- Hundreds of VAR coefficients to estimate – and they might be changing over time
- And large error covariances to estimates – and they might be changing over time
- ε_t is i.i.d. $N(0, \Sigma_t)$ and u_t is i.i.d. $N(0, Q_t)$.

Estimation Using Decay and Forgetting Factors

- Computational problem: recursively forecasting with TVP-VARs is hugely computationally demanding, even when VAR dimension is small (MCMC methods required)
- Forgetting factor approaches commonly used for estimating state space models in the past, when computing power was limited
- We use these (in a new context) to surmount computational burden
- Basic idea: if Σ_t and Q_t, known then computation vastly simplified
- Kalman filter and related methods for state space models can be used (no MCMC)
- For Σ_t use Matrix Wishart Discounting scheme
- Depends on discount factor, δ
- Provides not only point estimate of Σ_t but also posterior variance

- For Q_t use forgetting factor approximation
- Forgetting factor λ
- Details in paper
- Idea: δ and λ control degree of change in coefficients
- $\delta = \lambda = 1$ lead to homoskedastic, constant coefficient VAR
- $\delta < 1$ allows for volatility changes
- $\lambda < 1$ allows for VAR coeffs to change
- To preview our empirical results: volatility change is important ($\delta < 1$) but no evidence of coefficient change
- Main results set $\lambda = 1$ and $\delta < 1$

Minnesota Prior for VAR Coefficients

- Use Minnesota prior for β_0
- Preview empirical results: allowing for wide choice of shrinkage parameters is important
- We use up to 7 prior shrinkage parameters, $\gamma_1, .., \gamma_7$
- Different degree of shrinkage on: i) intercepts, ii) own lags, iii) other lags, iv) exogenous predictor 1, .., vii) exogenous predictor 4
- $\gamma_i = 0$ means coefficient (or block of coefficients) is set to zero
- γ_i larger means less prior shrinkage (data based estimation)

Model Selection Using Forgetting Factors

- So far have discussed one single model
- With many TVP regression models, Raftery et al (2010) develop methods for dynamic model selection (DMS) or dynamic model averaging (DMA)
- Different predictors can be selected at each point in time in a recursive forecasting exercise
- Basic idea: suppose $j = 1, .., J$ models.
- DMA/DMS calculate $\pi_{t|t-1,j}$: "probability that model j should be used for forecasting at time t, given information through time $t-1$"
- DMS: at each point in time forecast with model with highest value for $\pi_{t|t-1,j}$

Model Selection Using Forgetting Factors

- Raftery et al (2010) develop a fast recursive algorithm, similar to Kalman filter, using a forgetting factor for obtaining $\pi_{t|t-1,j}$.
- This involves a forgetting factor α which controls degree of model switching
- Interpretation of forgetting factor α
- Model j will receive more weight at time t if it has forecast well in the recent past
- Can show:

$$\pi_{t|t-1,j} = \prod_{i=1}^{t-1} \left[p_j \left(y_{t-i} = y_{t-i}^R | Data_{t-i-1} \right) \right]^{\alpha^i}$$

- Predictive likelihood of model j evaluated at realizations (y_{t-i}^R) given past data $(Data_{t-i-1})$

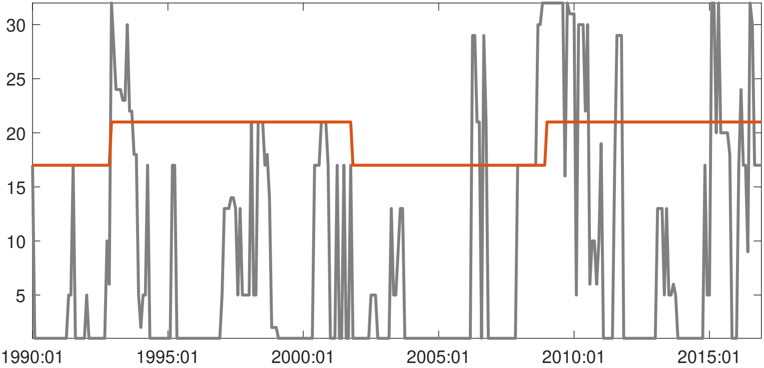
Model Selection Using Forgetting Factors

- Interpretation of "recent past" is controlled by the forgetting factor, α
- $\alpha = 0.99$: forecast performance five years ago receives 55% as much weight as forecast performance last period
- $\alpha = 0.95$: forecast performance five years ago receives only about 5% as much weight.
- $\alpha = 1$: can show $\pi_{t|t-1,k}$ is proportional to the marginal likelihood using data through time $t-1$ (standard BMA)
- Preview of empirical results: Choice of α very important
- We estimate α

Dynamic Model Learning (DML)

- We use DMS approach of Rafery et al (2010), but in a different way
- Raftery et al defines different models as using different predictors
- We define our models as involving different choices for γ, δ and λ
- Consider grid of values for them and use DMS to select optimal value at each point in time
- Also consider grid of values for α
- We find α and γ to be more important so main results set $\lambda = 1$, $\delta = 0.97$
- Wide grid on α (algorithm can select rate of model switching)
- Grids on γ_i for $i = 1, .., 7$ always include 0 (set coeffs or blocks of coeffs to zero)

- Focus of this paper is on economic value of our forecasting method
- But before those results, present some insight into what our approach is doing
- With long sample (no exogenous predictors) have 32 models
- Next figure:
- Our approach (grey line) lots of switching over time
- Setting $\alpha = 1$ (red line) very little switching


30
25
20
15
10
5
1990:01
1995:01
2000:01
2005:01
2010:01
2015:01

Empirical Results

- Model 1 is multivariate random walk (without drift but with stochastic volatility)
- Model 32 is unrestricted VAR with stochastic volatility
- Both extremes are chosen at some time or other
- Model 1 selected 55% of time (for short sample 30%)
- Our approach estimates α (degree of switching)
- Alternative of setting $\alpha = 1$ (conventional BMA) leads to much less switching (not supported by data)
- Which other models are selected? See paper

Forecasting and Portfolio Allocation

- Our approach provides one-month ahead predictive density
- Statistical evaluation of performance: log predictive likelihoods (LPL)
- Economic evaluation:
- Investor could use predictive density to build a portfolio and re-balance each month
- Φ^{TC} = performance fee after transactions cost
- = maximum (monthly) fee investor would be willing to pay to switch from portfolio based on random walks
- SR = Sharpe ratio
- SR^{TC} = Sharpe ratio after transactions costs

The Porfolio Allocation Problem

- Choose weights, $w_t = (w_{1,t}, ..., w_{9,t})'$ attached to each of 9 foreign bonds
- $1 - \sum_{i=1}^{9} w_{i,t}$ attached to the domestic bond (USD)
- Maximize expected portfolio return subject to target portfolio variance
- Solution (see paper) depends on predictive mean and predictive covariance matrix (produced by our algorithm)
- τ is transactions cost
- Following Della Corte et al, 2008, $\tau = 0.0008$ (results robust to alternative values)

Economic Evaluation

- Assume investor has quadratic utility
- Relative risk aversion, $\theta = 2$ (also tried $\theta = 6$)
- Economic value = average utility produced by allocating portfolio using our model relative to random walk
- Performance fee (Φ^{TC}) calculated using the economic value

List of Models Used in Forecasting Comparison

- Call our approach Dynamic Model Learning (DML)
- With long sample DML involves 9 exchange rates (discrete returns)
- With short sample DML = 9 exchange rates
- "DML with ALL" includes all exogenous predictors
- Also several special cases (nested in our approach)
- DML with one predictor entered at a time
- Variants which impose a particular value of α
- Exchange rates follow random walk (without drift but with stochastic volatility)
- Diebold-Mariano test of significant forecast improvement relative to random walk

Results Using Long Sample

- Next table gives results using long sample
- *DML* where α estimated, does best
- But only for economic evaluation criteria are results significantly better than random walk
- Note statistical criteria (*LPL*) very similar for all approaches
- *DML* where $\alpha = 0.8$ equally good
- But other results show that choice of α very important
- A priori, how would you know to set $\alpha = 0.8$? Better to estimate it
- Low values of α means lots of model switching

Results Using Long Sample

- Remember: DML chose multivariate random walk with stochastic volatility 55% of time
- But if we choose it 100% of time, then forecast performance deteriorates
- DML with cross lags deleted causes small drop in forecast performance
- DML with own lags deleted causes larger drop in forecast performance
- Overall: usually multivariate random walk with stochastic volatility is good model
- But sometimes must switch to richer VARs to achieve significant forecast improvements
- Need dynamic model learning

Evaluation of Forecasts: Long Sample				
	Φ^{TC}	SR	SR^{TC}	LPL
DML	485**	1.08**	0.92**	22.05
DML without cross lags	365**	0.82*	0.72*	21.86
DML without own lags	278	0.80	0.66	21.78
Random walk (stoch.vol.)	17	0.47	0.46	21.65
$DML(\alpha = 1)$	−255	0.35	0.19	21.65
$DML(\alpha = 0.9)$	238	0.78	0.65	21.88
$DML(\alpha = 0.8)$	485**	1.08**	0.92**	22.05
$DML(\alpha = 0.7)$	478**	1.07**	0.90**	22.06
$DML(\alpha = 0.5)$	409*	1.04**	0.84**	22.05

Results Using Short Sample

- Next table gives results using short sample
- Remember with short sample can include exogenous predictors
- Results point to usefulness of such predictors (especially UIP)
- Remember: 30% of time DML chooses random walk so UIP only important sometimes
- Performance fee (Φ^{TC}) results:
- *DML* (with no exogenous predictors) not significantly better than random walk
- *DML* with all exogenous predictors is significantly better than random walk
- Contrast with long sample: statistical criteria (*LPL*) also indicate significant improvements

Results Using Short Sample

- Dropping either own or cross lags leads to deterioration in economic performance
- *DML* where α estimated better than simply selecting α
- Some choices for α are bad
- With exception of *DML* with *UIP*, all results which are significant are only at 10% level

Evaluation of Forecasts: Short Sample				
	Φ^{TC}	SR	SR^{TC}	LPL
DML	327	101*	0.82	22.02*
DML with *OIL*	199	0.89	0.70	22.03*
DML with *UIP*	464*	1.12**	0.93**	22.01*
DML with *INTDIFF*	388*	1.06*	0.88*	22.02*
DML with *STOCKS*	368*	1.06*	0.88*	22.04*
DML with *ALL*	397*	1.02*	0.87*	22.04*
DML without cross lags	98	0.72	0.60	21.97
DML without own lags	200	0.86	0.79	21.78*
Random walk	5	0.54	0.53	21.72*
$DML(\alpha = 1)$	-427	0.34	0.11	21.69
$DML(\alpha = 0.9)$	98	0.77	0.60	21.96*
$DML(\alpha = 0.8)$	266	0.94	0.76	22.02*
$DML(\alpha = 0.7)$	327	1.01*	0.82*	22.02*
$DML(\alpha = 0.5)$	84	0.82	0.60	21.98

When Does DML Work Best?

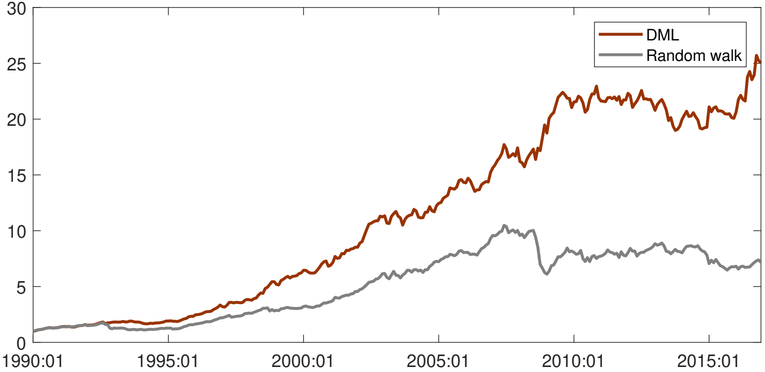
- Use DML (all regressors) and short sample and produce estimates of economic utility in each time period for each country
- Regress on some predictors:
- Shocks in FX volatility for the G10 countries ($\Delta FXVOL$)
- Disagreement between professional forecasters in FX ($FXDIS$)
- Shocks in volatility in equity markets (ΔVIX)
- Fitted regression line (t-statistics are in parentheses):

$$\Delta\widehat{Utility} = \underset{(1.51)}{0.0109} + \underset{(3.26)}{0.0031}\Delta FXVOL - \underset{(-0.20)}{0.00005}\Delta VIX + \underset{(1.43)}{0.0037}FXDIS$$

- It is in times of high FX volatility that our methods are working well

When Does DML Work Best?

- Evolution of wealth of two investors with $1 in 1990
- Investor 1 builds portfolio using DML
- Investor 2 builds portfolio using random walk
- Next figure plots evolution of wealth using these different investment dtrategies
- Shows when Investor 1 has biggest gains over Investor 2



DML
Random walk
30
25
20
15
10
5
0
1990:01
1995:01
2000:01
2005:01
2010:01
2015:01

Conclusion

- DML makes specification choices in dynamic fashion
- In each period, can have:
- Different degrees of prior shrinkage (blocks of coeffs removed from model for parsimony)
- Different time variation in parameters
- Different rate of model switching
- Using cross-section of exchange rates for G10 countries, investor using DML to build a portfolio would achieve appreciable economic gains