Large Time-Varying Parameter VARs

Gary Koop[1] Dimitris Korobilis[2]

[1]University of Strathclyde [2]University of Glasgow

May 30, 2012

Summary of Paper

- We extend large VAR literature to allow for time variation in parameters (VAR coefficients and error covariance matrix)
- Large TVP-VAR potentially over-parameterized, to deal with we do:
- Prior selection: degree of shrinkage selected automatically (and in a time-varying manner)
- Dynamic dimension selection (DDS): select dimension of TVP-VAR in time-varying manner
- Computational challenge over-come through use of forgetting factor methods
- Forgetting factors applied in a new way to allow for model switching
- Forecasting exercise using US data shows the approach works well

Large TVP-VARs

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

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

- if z_t is a vector containing an intercept and p lags of each of the M variables, then

$$Z_t = \begin{pmatrix} z_t' & 0 & \cdots & 0 \\ 0 & z_t' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z_t' \end{pmatrix}$$

- Note Z_t is $M \times k$ where $k = M(1 + pM)$
- VAR coefficients evolve according to:

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

- If $M = 25$, $p = 4$, then $k = 2525$
- Thousands of VAR coefficients to estimate – and they are all changing over time
- ε_t is i.i.d. $N(0, \Sigma_t)$ and u_t is i.i.d. $N(0, Q_t)$.

Forecasting with TVP-VARs Using 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)
- Replace Σ_t and Q_t by approximations
- For Σ_t use Exponentially Weighted Moving Average (EWMA) approximation (see paper for details)

Some Technical Details on Forgetting Factor treatment of Q

- Let $y^s = (y_1, .., y_s)'$ denote observations through time s.
- Kalman filter is standard tool for estimating state space models such as TVP-VAR
- Key steps in Kalman filtering involve the result:

$$\beta_{t-1}|y^{t-1} \sim N\left(\beta_{t-1|t-1}, V_{t-1|t-1}\right)$$

- Formulae for $\beta_{t-1|t-1}$ and $V_{t-1|t-1}$ are given in textbook sources.
- Kalman filtering then proceeds using:

$$\beta_t|y^{t-1} \sim N\left(\beta_{t|t-1}, V_{t|t-1}\right)$$

- where

$$V_{t|t-1} = V_{t-1|t-1} + Q_t$$

- This is only place where Q_t appears.

- Replace by:

$$V_{t|t-1} = \frac{1}{\lambda} V_{t-1|t-1}$$

- λ is called a forgetting factor, $0 < \lambda \leq 1$.
- Observations j periods in the past have weight λ^j in the estimation of β_t
- λ usually set to number slightly less than one.
- For quarterly macroeconomic data, $\lambda = 0.99$ implies observations five years ago receive approximately 80% as much weight as last period's observation.
- We also investigate estimating λ in a time varying manner.

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 model 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 $\pi_{t|t-1,j}$
- Raftery et al (2010) develop a fast recursive algorithm, similar to Kalman filter, using a forgetting factor for obtaining $\pi_{t|t-1,j}$.

- Interpretation of forgetting factor α
- Raftery's approach implies:

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

- $p_j \left(y_t | y^{t-1} \right)$ is the predictive likelihood (i.e. the predictive density for model j evaluated at y_t), produced by the Kalman filter
- Model j will receive more weight at time t if it has forecast well in the recent past
- Interpretation of "recent past" is controlled by the forgetting factor, α
- $\alpha = 0.99$: forecast performance five years ago receives 80% as much weight as forecast performance last period
- $\alpha = 0.95$: forecast performance five years ago receives only about 35% 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)

Model Selection Among Priors

- We use DMS approach of Rafery et al (2010), but in a different way
- Consider set of models defined by different priors
- Use popular Minnesota prior written as depending on one shrinkage parameter γ
- Consider grid of values for γ and use DMS to select optimal value at each point in time

Model Selection Among TVP-VARs of Different Dimension

- Use DMS approach over three models: a small, medium and large TVP-VAR.
- Small: contains variables we want to forecast (GDP growth, inflation and interest rates)
- Medium: variables in small model plus four others suggested by DSGE literature
- Large: variables in medium model plus 18 others often used to forecast inflation or output growth
- Note: $p_j \left(y_{t-i} | y^{t-i-1} \right)$, plays the key role in DMS.
- We use predictive likelihood for the 3 variables in the small model (common to all approaches)

Empirical Results: Data and Modelling Issues

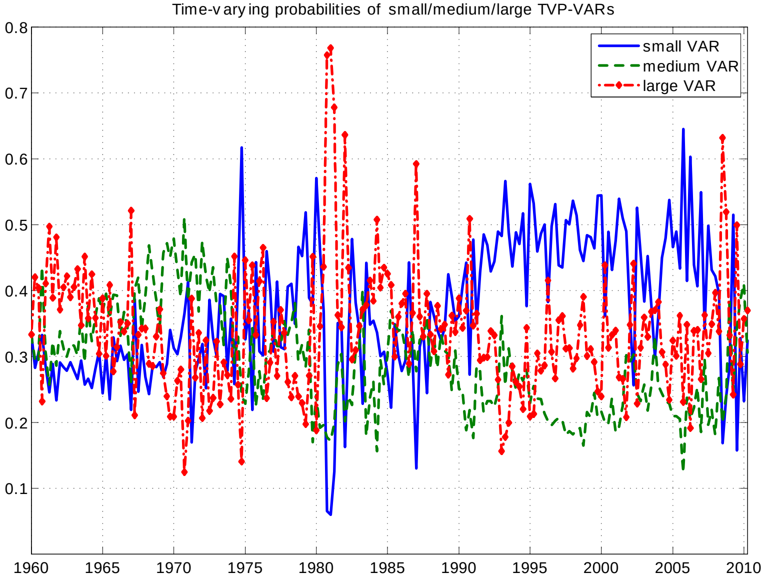
- 25 major quarterly US macroeconomic variables, 1959:Q1 to 2010:Q2.
- Following, e.g., Stock and Watson (2008) and recommendations in Carriero, Clark and Marcellino (2011) we transform all variables to stationarity.
- We use a lag length of 4.
- Time-variation in the VAR coefficients: $\lambda = 0.99$.
- Degree of model switching: $\alpha = 0.99$.
- EWMA discount factor, controls the volatility, $\kappa = 0.96$.

Other Models Used for Comparison

- TVP-VARs of each dimension, with no DDS being done.
- Time-varying forgetting factor versions of the TVP-VARs.
- VARs of each dimension
- Homoskedastic versions of each VAR.
- Random walk forecasts (labelled RW)
- A small VAR estimated using OLS methods

Evidence of Model Change

- Next figure shows probabilities DDS produces for TVP-VARs of different dimensions
- DDS will choose model with highest probability
- Lots of evidence for dimension switching
- Small TVP-VAR used to forecast mostly from 1990-2007
- Large TVP-VAR typically used in 1980s
- Medium TVP-VAR in early 1970s
- Similar evidence of model switching for shrinkage parameter (see paper)



Time-varying probabilities of small/medium/large TVP-VARs
small VAR
medium VAR
large VAR
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010

Forecast Comparison

- Iterated forecasts for horizons of up to two years ($h = 1, .., 8$)
- Forecast evaluation period of 1970Q1 through 2010Q2.
- Note: with iterated forecasts for $h > 1$ predictive simulation is required
- We do this in two ways.
- 1. VAR coefficients which hold at T used to forecast at $T + h$ ($\beta_{T+h} = \beta_T$)
- 2. $\beta_{T+h} \sim RW$ simulates from random walk state equation to produce draws of β_{T+h}.
- Both ways provide us with β_{T+h}, we simulate draws of y_{T+h} conditional on β_{T+h} to approximate the predictive density.
- Measures of forecast performance:
- Mean squared forecast errors (MSFEs) — evaluate quality of point forecasts
- Sums of log predictive likelihoods: use the joint predictive likelihood for these three variables – evaluate quality of entire predictive distribution

Summary of Results for Predictive Likelihoods

- MSFE results (see paper)
- MSFE story: TVP-VAR-DDS is forecasting better than simple benchmarks or VARs/TVP-VARs of fixed dimension
- Table 4 presents sums of log predictive likelihoods for a specific model minus that of TVP-VAR-DDS
- Negative numbers indicate our approach is forecasting better
- Almost all of these numbers are negative (reinforces story told by MSFEs)
- At $h = 1$, TVP-VAR-DDS forecasts best by considerable margin and at other horizons beats other TVP-VAR approaches.

- One difference between predictive likelihood and MSFE results:
- Importance of allowing for heteroskedastic errors is more evident
- It is key in getting the shape of the predictive density correct
- Heteroskedastic VAR exhibits best forecast performance at some horizons for some variables.
- But dimensionality of best heteroskedastic VAR differs across horizons (sometimes small VAR best, other times large)
- Message: even when researcher is using a VAR (instead of a TVP-VAR), DDS still might be useful where there is uncertainty over dimension of VAR.

Table 4a: Relative Predictive Likelihoods, Total (all 3 variables)				
	$h = 1$	$h = 2$	$h = 4$	$h = 8$
Full model				
TVP-VAR-DDS, $\lambda = 0.99, \beta_{T+h} = \beta_T$	0.84	0.91	4.03	4.11
TVP-VAR-DDS, $\lambda = 0.99, \beta_{T+h} \sim RW$	0.00	0.00	0.00	0.00
Small VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	-6.71	4.62	-2.72	0.68
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	-7.47	2.15	-3.72	-3.63
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	-5.95	4.84	-2.56	-3.32
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	-4.77	3.70	-0.68	3.36
VAR, heteroskedastic	-6.18	6.86	1.57	9.11
VAR, homoskedastic	-47.44	-29.97	-22.87	-15.93
Medium VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	-23.55	0.79	2.84	9.27
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	-30.24	-6.10	0.05	10.68
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	-23.22	-0.09	-0.54	9.80
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	-20.69	0.68	1.62	4.87
VAR, heteroskedastic	-20.89	1.08	8.39	14.52
VAR, homoskedastic	-58.28	-31.86	-21.09	-10.65

Table 4b: Relative Predictive Likelihoods, Total (all 3 variables)				
	$h=1$	$h=2$	$h=4$	$h=8$
LARGE VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	-18.16	-7.81	-1.32	8.33
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	-21.96	-12.99	-10.61	-2.82
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	-16.14	-8.25	-2.45	2.93
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	-16.24	-5.20	-0.41	1.82
VAR, heteroskedastic	-17.30	-1.63	8.46	13.24
VAR, homoskedastic	-50.33	-37.35	-28.60	-20.50
BENCHMARK MODELS				
RW	-	-	-	-
Small VAR OLS	-52.94	-40.42	-52.48	-49.35

Conclusions

- We have developed method for forecasting with large TVP-VARs using forgetting factors.
- Forgetting factors useful in 3 ways
- 1. Computationally feasible forecasting within a single TVP-VAR model.
- 2. Dynamic prior selection where degree of shrinkage estimated in a time-varying fashion.
- 3. Dynamic dimension selection : TVP-VAR dimension may change over time.
- Empirical work: forecasting US inflation, GDP growth and interest rates
- Small, medium and large TVP-VARs and VARs
- We find moderate improvements in forecast performance over other VAR or TVP-VAR approaches.