TVP-VARs

Bank of Korea Global Initiative Program

September 2014

Introduction

- Why TVP-VARs?
- Example: U.S. monetary policy
- was the high inflation and slow growth of the 1970s were due to bad policy or bad luck?
- Some have argued that the way the Fed reacted to inflation has changed over time
- After 1980, Fed became more aggressive in fighting inflation pressures than before
- This is the "bad policy" story (change in the monetary policy transmission mechanism)
- This story depends on having VAR coefficients different in the 1970s than subsequently.

- Others think that variance of the exogenous shocks hitting economy has changed over time
- Perhaps this may explain apparent changes in monetary policy.
- This is the "bad luck" story (i.e. 1970s volatility was high, adverse shocks hit economy, whereas later policymakers had the good fortune of the Great Moderation of the business cycle – at least until 2008)
- This motivates need for multivariate stochastic volatility to VAR models
- Cannot check whether volatility has been changing with a homoskedastic model

- Most macroeconomic applications of interest involve several variables (so need multivariate model like VAR)
- Also need VAR coefficients changing
- Also need multivariate stochastic volatility
- TVP-VARs are most popular models with such features
- But other exist (Markov-switching VARs, Vector Floor and Ceiling Model, etc.)

Homoskedastic TVP-VARs

- Begin by assuming $\Sigma_t = \Sigma$
- Remember VAR notation: y_t is $M \times 1$ vector, Z_t is $M \times k$ matrix (defined so as to allow for a VAR with different lagged dependent and exogenous variables in each equation).
- TVP-VAR:

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

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

- ε_t is i.i.d. $N(0, \Sigma)$ and u_t is i.i.d. $N(0, Q)$.
- ε_t and u_s are independent of one another for all s and t.

- Bayesian inference in this model?
- Already done: this is just the Normal linear state space model of the last lecture.
- MCMC algorithm of standard form (e.g. Carter and Kohn, 1994).
- But let us see how it works in practice in our empirical application
- Follow Primiceri (2005)

Illustration of Bayesian TVP-VAR Methods

- Same quarterly US data set from 1953Q1 to 2006Q3 as was used to illustrate VAR methods
- Three variables: Inflation rate $\Delta\pi_t$, the unemployment rate u_t and the interest rate r_t
- VAR lag length is 2.
- Training sample prior: prior hyperparameters are set to OLS quantities calculating using an initial part of the data
- Our training sample contains 40 observations.
- Data through 1962Q4 used to choose prior hyperparameter values, then Bayesian estimation uses data beginning in 1963Q1.

- β_{OLS} is OLS estimate of VAR coefficients in constant-coefficient VAR using training sample
- $V\left(\beta_{OLS}\right)$ is estimated covariance of β_{OLS}.
- Prior for β_0:
$$\beta_0 \sim N\left(\beta_{OLS}, 4 \cdot V\left(\beta_{OLS}\right)\right)$$
- Prior for Σ^{-1} Wishart prior with $\underline{\nu} = M + 1, \underline{S} = I$
- Prior for Q^{-1} Wishart prior with $\underline{\nu}_Q = 40, \underline{Q} = 0.0001 \cdot 40 \cdot V\left(\beta_{OLS}\right)$

- With TVP-VAR we have different set of VAR coefficients in every time period
- So different impulse responses in every time period.
- Figure 1 presents impulse responses to a monetary policy shock in three time periods: 1975Q1, 1981Q3 and 1996Q1.
- Impulse responses defined in same way as we did for VAR
- Posterior median is solid line and dotted lines are 10^{th} and 90^{th} percentiles.

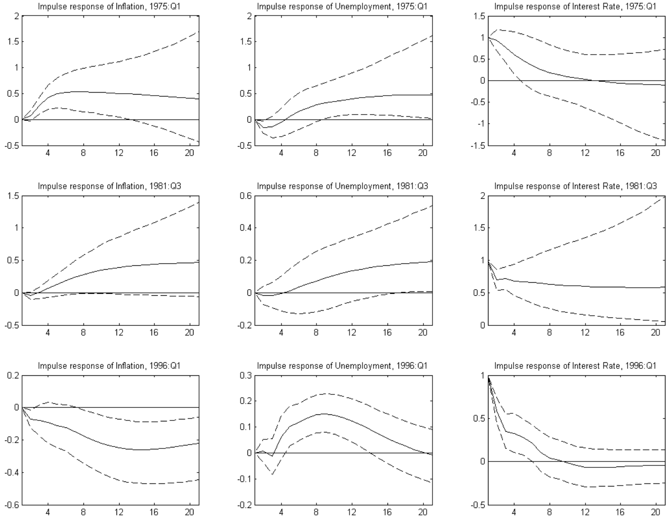


Figure 1: Impulse responses at different times

Learning Bayesian Computation

- Remember: this course has a lecture format
- But I have made up some computer question sheets that you you may wish to work through on your own
- Computer session 4 (on the course website) has questions relating to the empirical illustration above
- MATLAB computer code which answers the questions is also available on the website

Combining other Priors with the TVP Prior

- Often Bayesian TVP-VARs work very well in practice.
- In some case the basic TVP-VAR does not work as well, due to over-parameterization problems.
- Previously, we noted worries about proliferation of parameters in VARs, which led to use of priors such as the Minnesota prior or the SSVS prior.
- With many parameters and short macroeconomic time series, it can be hard to obtain precise estimates of coefficients.
- Risk of over-fitting
- Priors which exhibit shrinkage of various sorts can help mitigate these problems.
- With TVP-VAR proliferation of parameters problems is even more severe.
- Hierarchical prior of state equation is big help, but may want more in some cases.

Combining TVP Prior with Minnesota Prior

- E.g. Ballabriga, Sebastian and Valles (1999, JIE), Canova and Ciccarelli (2004, JOE), and Canova (2007, book)
- Replace TVP-VAR state equation by

$$\beta_{t+1} = A_0\beta_t + (I - A_0)\,\overline{\beta}_0 + u_t$$

- u_t is i.i.d. $N(0, Q)$
- A_0, $\overline{\beta}_0$ and Q can be unknown parameters or set to known values
- E.g. Canova (2007) sets $\overline{\beta}_0$ and Q to have forms based on the Minnesota prior and sets $A_0 = cI$ where c is a scalar.
- Note if $c = 1$, then $E\left(\beta_{t+1}\right) = E\left(\beta_t\right)$ (as in TVP-VAR)
- If $c = 0$ then $E\left(\beta_{t+1}\right) = \overline{\beta}_0$ (as in Minnesota prior)
- Q based on prior covariance of Minnesota prior
- c can either be treated as an unknown parameter or a value can be selected for it.

Combining TVP Prior with SSVS Prior

- Same setup as preceding slide
- Set $\overline{\beta}_0 = 0$.
- Let $a_0 = vec\left(A_0\right)$
- Use SSVS prior for a_0
- a_{0j} (the j^{th} element of a_0) has prior:

$$a_{0j}|\gamma_j \sim \left(1 - \gamma_j\right) N\left(0, \kappa_{0j}^2\right) + \gamma_j N\left(0, \kappa_{1j}^2\right)$$

- as before, γ_j is dummy variable
- κ_{0j}^2 is very small (so that a_{0j} is constrained to be virtually zero)
- κ_{1j}^2 is large (so that a_{0j} is relatively unconstrained).
- Property: with probability γ_j, a_{0j} is evolving according to a random walk in the usual TVP fashion
- With probability $\left(1 - \gamma_j\right)$, $a_{0j} \approx 0$

MCMC Algorithm

- I will not provide complete details, but note only:
- These are Normal linear state space models so standard algorithms (e.g. Carter and Kohn) can draw β^T
- For TVP+Minnesota prior this is enough (other parameters fixed)
- For TVP+SSVS simple to adapt MCMC algorithm for SSVS with VAR

Adding Another Layer to the Prior Hierarchy

- Another approach used by Chib and Greenberg (1995, JOE) for SUR model
- Adapted for VARs by, e.g., Ciccarelli and Rebucci (2002)
- $$\begin{aligned} y_t &= Z_t \beta_t + \varepsilon_t \\ \beta_{t+1} &= A_0 \theta_{t+1} + u_t \\ \theta_{t+1} &= \theta_t + \eta_t \end{aligned}$$
- all assumptions as for TVP-VAR, plus η_t is i.i.d. $N(0, R)$
- Slightly more general that previous Normal linear state space model, but very similar MCMC (so will not discuss MCMC)

Adding Another Layer to the Prior Hierarchy

- Why might this generalization be useful?
- A_0 can be chosen to reflect some other prior information
- E.g. SSVS prior as above
- E.g. Ciccarelli and Rebucci (2002) is panel VAR application
- G countries and, for each country, k_G explanatory variables exist with time-varying coefficients.
- They set

$$A_0 = \iota_G \otimes I_{k_G}$$

- Implies time-varying component in each coefficient which is common to all countries
- Parsimony: θ_t is of dimension k_G whereas β_t is of dimension $k_G \times G$.

Imposing Inequality Restrictions on the VAR Coefficients

- Another way of ensuring shrinkage
- E.g. restrict β_t to be non-explosive (i.e. roots of the VAR polynomial defined by β_t lie outside the unit circle)
- Sometimes (given over-fitting and imprecise estimates) can get posterior weight in explosive region
- Even small amount of posterior probability in explosive regions for β_t can lead to impulse responses or forecasts which have counter-intuitively large posterior means or standard deviations.
- Koop and Potter (2009, on my website) discusses how to do this. I will not present details, but outline basic idea

- With unrestricted TVP-VAR, took draws $p\left(\beta^T | y^T, \Sigma, Q\right)$ using MCMC methods for Normal linear state space models
- One method to impose inequality restrictions involves:
- Draw β^T in the unrestricted VAR. If *any* drawn β_t violates the inequality restriction then the *entire* vector β^T is rejected.
- Problem: this algorithm can get stuck, rejecting virtually every β^T (all you need is a single drawn β_t to violate inequality and entire β^T is rejected)
- Note: algorithms like Carter and Kohn are "multi-move algorithms" (draw β^T all at same time).
- Alternative is "single move algorithm": drawing β_t for $t = 1, .., T$ one at a time from $p\left(\beta_t | y^T, \Sigma, Q, \beta_{-t}\right)$ where $\beta_{-t} = \left(\beta_1', .., \beta_{t-1}, \beta_{t+1}, .., \beta_T'\right)'$

- Koop and Potter (2009) suggest using single move algorithm
- Reject β_t only (not β^T) if it violates inequality restriction
- Usually multi-move algorithms are better than single-move algorithms since latter can be slow to mix.
- I.e. they produce highly correlated series of draws which means that, relative to multi-move algorithms, more draws must be taken to achieve a desired level of accuracy.
- But if multi-move algorithm gets stuck, single move might be better.

Dynamic Mixture Models

- Remember: Normal linear state space model depends on so-called system matrices, Z_t, Q_t, T_t, W_t and Σ_t.
- Suppose some or all of them depend on an $s \times 1$ vector $\widetilde{K}_t$
- Suppose $\widetilde{K}_t$ is Markov random variable (i.e. $p\left(\widetilde{K}_t | \widetilde{K}_{t-1}, .., \widetilde{K}_1\right) = p\left(\widetilde{K}_t | \widetilde{K}_{t-1}\right)$ or independent over t
- Particularly simple if $\widetilde{K}_t$ is a discrete random variable.
- Result is called a dynamic mixture model
- Gerlach, Carter and Kohn (2000, JASA) have an efficient MCMC algorithm

- Why are dynamic mixture models useful in empirical macroeconomics?
- E.g. TVP-VAR:

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

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

- ε_t is i.i.d. $N(0, \Sigma)$
- BUT: u_t is i.i.d. $N\left(0, \widetilde{K}_t Q\right)$.
- Let $\widetilde{K}_t \in \{0, 1\}$ with hierarchical prior:

$$p\left(\widetilde{K}_t = 1\right) = q.$$
$$p\left(\widetilde{K}_t = 0\right) = 1 - q$$

- where q is an unknown parameter.

- Property:
- If $\widetilde{K}_t = 1$ then usual TVP-VAR:

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

- If $\widetilde{K}_t = 0$ then VAR coefficients do not change at time t:

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

- Parsimony.
- This model can have flexibility of TVP-VAR if the data warrant it (i.e. can select $\widetilde{K}_t = 1$ for $t = 1, .., T$).
- But can also select a much more parsimonious representation.
- An extreme case: if $\widetilde{K}_t = 0$ for $t = 1, .., T$ then back to VAR without time-varying parameters.

- I will not present details of MCMC algorithm since it is becoming a standard one
- See also the Matlab code on my website
- Dynamic mixture models used to model structural breaks, outliers, nonlinearities, etc.
- E.g. Giordani, Kohn and van Dijk (2007, JoE).

TVP-VARs with Stochastic Volatility

- In empirical work, you will usually want to add multivariate stochastic volatility to the TVP-VAR
- But this can be dealt with quickly, since the appropriate algorithms were described in the lecture on State Space Modelling
- Remember, in particular, the approaches of Cogley and Sargent (2005) and Primiceri (2005).
- MCMC: need only add another block to our algorithm to draw Σ_t for $t = 1, .., T$.
- Homoskedastic TVP-VAR MCMC: $p\left(Q^{-1}|y^T, \beta^T\right)$, $p\left(\beta^T|y^T, \Sigma, Q\right)$ and $p\left(\Sigma^{-1}|y^T, \beta^T\right)$
- Heteroskedastic TVP-VAR MCMC: $p\left(Q^{-1}|y^T, \beta^T\right)$, $p\left(\beta^T|y^T, \Sigma_1, .., \Sigma_T, Q\right)$ and $p\left(\Sigma_1^{-1}, .., \Sigma_T^{-1}|y^T, \beta^T\right)$

Empirical Illustration of Bayesian Inference in TVP-VARs with Stochastic Volatility

- Continue same illustration as before.
- All details as for homoskedastic TVP–VAR
- Plus allow for multivariate stochastic volatility as in Primiceri (2005).
- Priors as in Primiceri
- Can present empirical features of interest such as impulse responses
- But (for brevity) just present volatility information
- Figure 2: time-varying standard deviations of the errors in the three equations (i.e. the posterior means of the square roots of the diagonal element of Σ_t)
- If time permits, I will show empirical results from dynamic mixture version of model from my paper with Leon-Gonzalez and Strachan (working paper version available on my website)

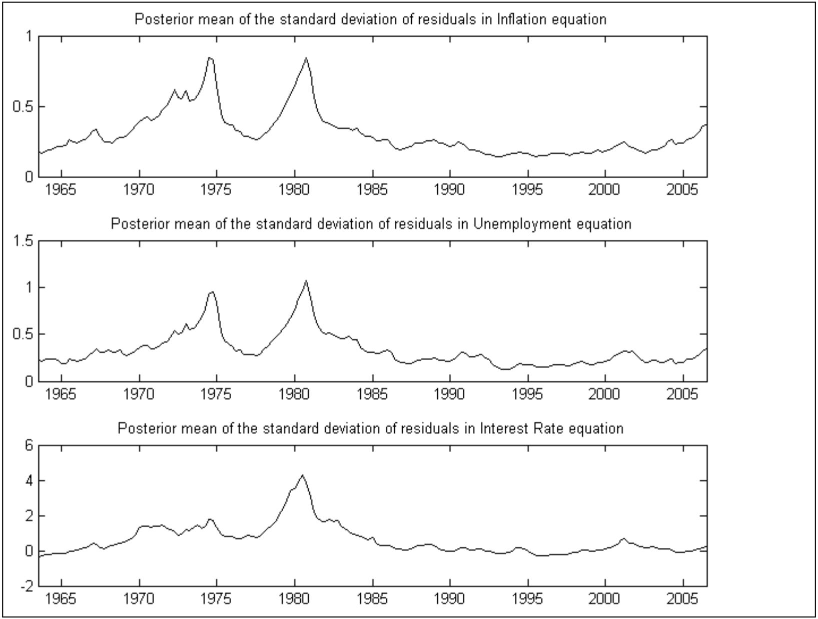


Figure 2: Volatilities in the 3 Equations

Empirical Application Using Large TVP-VARs

- With VARs, we saw how there was a growing interest in large VARs
- But large VARs over-parameterized, need for lots of prior shrinkage
- With TVP-VARs such problems will be magnified when we move to larger models
- Computational challenges are significant with large VARs, but may be insurmountable with large TVP-VARs
- The paper Koop and Korobilis (2013, JoE), "Large Time-Varying Parameter VARs" is an attempt to develop methods for over-coming these challenges
- This application is based on our paper

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:

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

- 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 computationally demanding, even when dimension is small (MCMC methods required)
- In previous lecture introduced DMA methods involving forgetting factors were discussed
- 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 Q_t use forgetting factor approximation described previously
- λ is forgetting factor
- Replace Σ_t and Q_t by approximations
- For Σ_t use Exponentially Weighted Moving Average (EWMA) approximation (see paper for details)
- κ is decay factor in EWMA

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
- Reminder of 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 $\pi_{t|t-1,j}$.
- Forgetting factor is α
- Properties of forgetting factor approaches discussed in last lecture

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
- This is the simple variant of the Minnesota prior discussed in empirical illustration in lecture 2
- Consider grid of values for γ and use DMS to select optimal value at each point in time
- Some details on next slide

Model Selection Among Priors

- Minnesota prior shrinks coefficients towards a prior mean (in our case zero)
- Prior variance controls amount of shrinkage
- $\underline{V}_i$ contains prior variances for coefficients in equation i
- We set:

$$\underline{V}_i = \begin{cases} \frac{\gamma}{r^2} \text{ for coefficients on lag } r \text{ for } r = 1, .., p \\ \underline{a} \text{ for the intercepts} \end{cases}$$

- $\underline{a}$ set to large number (noninformative prior)
- shrinkage gets tighter for as lag length increases
- γ controls shrinkage.
- We select over grid: $\left[10^{-10}, 10^{-5}, 0.001, 0.005, 0.01, 0.05, 0.1\right]$.

Model Selection Among TVP-VARs of Different Dimension

- We also 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 also use predictive likelihood for the 3 variables in the small model (common to all approaches)

Summary

- We are working with TVP-VARs
- Set of models defined by different degrees of shrinkage and different dimensionality
- DMS allows us to select the best forecasting model at each point in time
- So shrinkage may change over time
- Model dimensionality may change over time
- We call the latter DDS = dynamic dimension selection

Empirical Results: Data and Modelling Issues

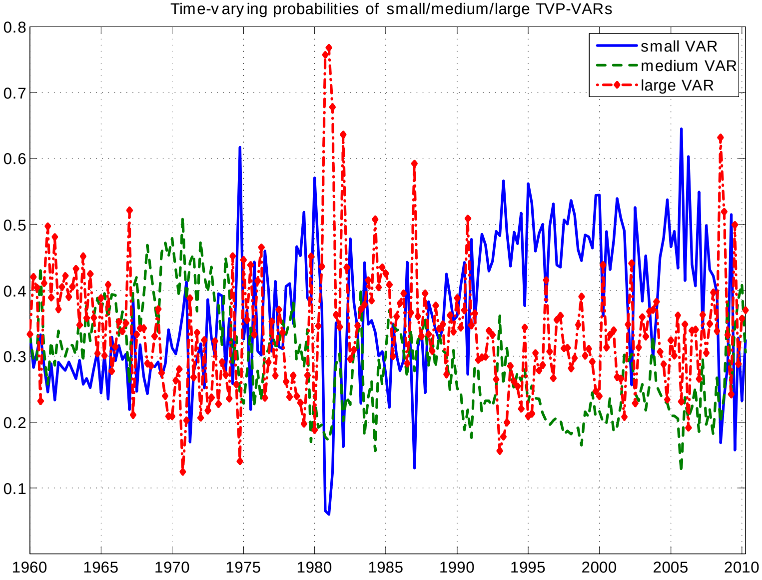
- 25 major quarterly US macroeconomic variables, 1959:Q1 to 2010:Q2.
- Following, e.g., Stock and Watson (2008) and recommendations in Carriero, Clark and Marcellino (2011) we transform all variables to stationarity.
- We use a lag length of 4.
- Time-variation in the VAR coefficients: $\lambda = 0.99$.
- 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
- etc.
- 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 MSFEs (if time permits)

- Tables 1, 2 and 3 compare MSFEs for forecasting GDP growth, inflation and interest rates, respectively
- MSFEs presented relative to the TVP-VAR-DDS approach which simulates β_{T+h} from the random walk state equation.
- Hence, numbers greater than one indicate our approach is forecast better
- Overall pattern: almost all numbers in tables are greater than one (particularly true for inflation and GDP)
- TVP-VAR-DDS is forecasting much better than our most simple benchmarks: random walk forecasts and forecasts from a small VAR estimated using OLS methods.
- Allowing for TVP-VAR dimensionality to change almost always works better than simply working with a fixed dimension

Table 1a: Relative Mean Squared Forecast Errors, GDP equation

	$h=1$	$h=2$	$h=4$	$h=8$
FULL MODEL				
TVP-VAR-DDS, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.00	1.02	1.03	0.99
TVP-VAR-DDS, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.00	1.00	1.00	1.00
SMALL VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.04	0.95	1.00	1.02
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	1.03	0.92	1.03	1.04
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.05	0.95	1.03	1.02
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	1.04	0.95	1.02	1.01
VAR, heteroskedastic	1.04	0.94	1.03	1.04
VAR, homoskedastic	1.09	1.01	1.01	1.04
MEDIUM VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.09	0.99	1.04	1.07
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	1.09	0.99	1.04	1.06
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.10	1.00	1.07	1.05
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	1.05	1.00	1.04	1.10
VAR, heteroskedastic	1.10	1.00	1.05	1.10
VAR, homoskedastic	1.08	1.02	1.08	1.08

Table 1b: Relative Mean Squared Forecast Errors, GDP equation

	$h = 1$	$h = 2$	$h = 4$	$h = 8$
Large VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.03	1.04	1.06	1.10
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	1.04	1.06	1.10	1.11
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.02	1.05	1.06	1.09
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	1.05	1.10	1.05	1.11
VAR, heteroskedastic	1.09	1.12	1.11	1.13
VAR, homoskedastic	1.02	1.05	1.04	1.05
Benchmark Models				
RW	1.59	1.71	1.97	2.22
Small VAR OLS	1.19	1.13	1.29	1.29

Table 2a: Relative Mean Squared Forecast Errors, Inflation equation

	$h=1$	$h=2$	$h=4$	$h=8$
FULL MODEL				
TVP-VAR-DDS, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.02	0.99	1.00	1.00
TVP-VAR-DDS, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.00	1.00	1.00	1.00
SMALL VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.04	1.05	1.06	1.04
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	1.04	1.06	1.06	1.03
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.03	1.06	1.06	1.04
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	1.03	1.07	1.03	1.06
VAR, heteroskedastic	1.02	1.04	1.01	1.05
VAR, homoskedastic	1.05	1.08	1.02	1.06
MEDIUM VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.08	1.06	1.01	1.05
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	1.12	1.07	0.99	1.07
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.08	1.05	1.01	1.04
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	1.07	1.05	1.02	1.07
VAR, heteroskedastic	1.07	1.06	1.00	1.07
VAR, homoskedastic	1.11	1.10	1.03	1.09

Table 2b: Relative Mean Squared Forecast Errors, Inflation equation				
	$h=1$	$h=2$	$h=4$	$h=8$
LARGE VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.01	1.02	0.95	1.04
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	1.01	1.04	0.95	1.04
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.01	1.03	0.95	1.02
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	1.03	1.01	0.96	1.05
VAR, heteroskedastic	1.05	1.03	0.95	1.04
VAR, homoskedastic	1.05	1.05	0.96	1.07
BENCHMARK MODELS				
RW	3.26	2.71	2.07	1.74
Small VAR OLS	1.09	1.23	1.14	1.18

Table 3a: Relative Mean Squared Forecast Errors, Interest Rate equation

	$h=1$	$h=2$	$h=4$	$h=8$
Full model				
TVP-VAR-DDS, $\lambda=0.99, \beta_{T+h}=\beta_T$	1.03	1.00	1.00	0.99
TVP-VAR-DDS, $\lambda=0.99, \beta_{T+h}\sim RW$	1.00	1.00	1.00	1.00
Small VAR				
TVP-VAR, $\lambda=0.99, \beta_{T+h}=\beta_T$	1.16	1.02	1.19	1.11
TVP-VAR, $\lambda=\lambda_t, \beta_{T+h}=\beta_T$	1.18	0.99	1.12	1.07
TVP-VAR, $\lambda=0.99, \beta_{T+h}\sim RW$	1.16	1.00	1.20	1.11
TVP-VAR, $\lambda=\lambda_t, \beta_{T+h}\sim RW$	1.19	1.01	1.16	1.08
VAR, heteroskedastic	1.19	1.00	1.09	1.01
VAR, homoskedastic	1.25	1.10	1.11	1.03
Medium VAR				
TVP-VAR, $\lambda=0.99, \beta_{T+h}=\beta_T$	1.18	1.01	1.06	0.97
TVP-VAR, $\lambda=\lambda_t, \beta_{T+h}=\beta_T$	1.19	1.03	1.06	0.98
TVP-VAR, $\lambda=0.99, \beta_{T+h}\sim RW$	1.20	1.01	1.06	0.98
TVP-VAR, $\lambda=\lambda_t, \beta_{T+h}\sim RW$	1.19	0.98	1.04	0.98
VAR, heteroskedastic	1.17	0.97	1.02	0.96
VAR, homoskedastic	1.25	1.06	1.03	0.98

Table 3b: Relative Mean Squared Forecast Errors, Interest Rate equation

	$h=1$	$h=2$	$h=4$	$h=8$
LARGE VAR				
TVP-VAR, $\lambda = 0.99, \beta_{T+h} = \beta_T$	1.07	0.94	0.96	0.92
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} = \beta_T$	1.06	0.97	0.98	0.92
TVP-VAR, $\lambda = 0.99, \beta_{T+h} \sim RW$	1.05	0.94	0.97	0.91
TVP-VAR, $\lambda = \lambda_t, \beta_{T+h} \sim RW$	1.07	0.93	0.97	0.91
VAR, heteroskedastic	1.07	0.95	0.97	0.91
VAR, homoskedastic	1.13	0.98	0.99	0.92
BENCHMARK MODELS				
RW	1.91	2.16	1.87	1.93
Small VAR OLS	1.76	1.47	2.11	2.03

Summary of Results for Predictive Likelihoods

- 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

Summary of Results for Predictive Likelihoods

- 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.
- If time permits, I will present my recent paper on this topic

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.

Summary

- TVP-VARs are useful for the empirical macroeconomists since they:
- are multivariate
- allow for VAR coefficients to change
- allow for error variances to change
- They are state space models so Bayesian inference can use familar MCMC algorithms developed for state space models.
- They can be over-parameterized so care should be taken with priors.
- When working with large TVP-VARs computation can also be a major worry, but approximate methods (involving forgetting factors) seem to work well