Large VARs with Time-Varying parameters: a nonparametric approach

George Kapetanios[1], Massimiliano Marcellino[2] and Fabrizio Venditti[3]

[1]King's College and Bank of England
[2]Bocconi University/CEPR/IGIER [3]Bank of Italy

ECB Workshop on 'Forecasting Techniques: Forecast Uncertainty and Macroeconomic Indicators', Frankfurt, 3-4 June, 2016

Motivation

- In the last 10/15 years intensive work in applied macro-econometrics in two areas:
  - Large models
  - Time varying parameter models
- Large models have become the dominant framework in forecasting
- TVP models, especially TVP VARs, extensively used for policy analysis
- Interest in connecting the two fields: Large TVP models

Large models

Problem of large models: parameter proliferation/overfitting

Shrink the information

- Dynamic factor models
- Factor augmented VARs

Shrink the parameters

- Penalized regressions (Ridge/Lars/Lasso)
- Large Bayesian VARs with tight Priors

De Mol, Giannone, Reichlin (2008) show that there is indeed a connection between the two approaches

TVP-VARs

Parametric

Bayesian

- small TVP-BVARs (Cogley/Sargent/Primiceri)

Score-driven

- AR with T.V.P. and heteroschedastic errors (Delle Monache and Petrella, 2015), now extending to VARs

Non-Parametric

- Giraitis, L., G. Kapetanios, and T. Yates (2014): AR with T.V.P. and heteroschedastic errors
- Giraitis, L., G. Kapetanios, and T. Yates (2014, GKY henceforth): VAR with T.V.P. and heteroschedastic errors

TVP-VARs in data rich environment

Parametric

Koop and Korobilis (2013) large Bayesian TVP-VAR

Non-parametric

THIS PAPER!

Roadmap of the presentation

Methodology

- The GKY estimator
- Taking the GKY estimator to large data: stochastic constraints
- Special cases: TVP Ridge estimator, TVP VAR with Litterman type constraints

Applications

- Forecasting in a data rich environment
- Small sample performance and a comparison with the parametric estimator
- Time-varying effects of oil price shocks on US industrial production

The GKY estimator 1

Setup of the problem:

$$\underset{1\times n}{y_t'} = x_t'\Theta_t + \underset{1\times n}{u_t'}, \quad t = 1, ..., T$$
$$\underset{1\times k}{x_t'} = [y_{t-1}', y_{t-2,...,}' y_{t-p}', 1]$$
$$\underset{k\times n}{\Theta_t} = [\Theta_{t,1}', \Theta_{t,2}^{'}, ..., \Theta_{t,p}', A_t']'$$

where $k = (np+1)$. At each time t you have nk parameters to estimate. Applying the *vec* operator to both sides we obtain:

$$\underset{n\times 1}{y_t} = \underset{n\times nk}{(I_n \otimes x_t')} \underset{nk\times 1}{\beta_t} + \underset{n\times 1}{u_t},$$

where $\beta_t = vec(\Theta_t)$.

The GKY estimator 2

GKY propose the following estimator:

$$\underset{nk\times 1}{\widehat{\beta}_t^{GKY}} = \underbrace{I_n \otimes \underbrace{\left[\sum_{j=1}^{T} w_{j,t}^{H}\left(x_j' x_j\right)\right]^{-1}}_{k\times k}}_{nk\times nk} \underbrace{\left[vec \sum_{j=1}^{T} w_{j,t}^{H}\left(x_j' y_j'\right)\right]}_{nk\times 1}$$

where $w_{j,t}^{H}$ is a kernel (weight) function that discounts sample moments
It depends on

- distance between t (time of interest for the parameters) and j (time of reference of the sample moment)
- normalized by H (the bandwidth). The smaller the bandwidth the stronger the discounting

kernel

Gaussian kernel

$$w_{j,t}^{H} = \frac{K_{j,t}(H)}{\sum_{j=1}^{T} K_{j,t}(H)} \quad \text{where}$$

$$K_{j,t}(H) = (1/\sqrt{2\pi}) exp \left[-\frac{1}{2} \left(\frac{j-t}{H} \right)^2 \right]$$

When forecasting

$$K_{j,t}(H) = (1/\sqrt{2\pi}) exp \left[-\frac{1}{2} \left(\frac{j-t}{H} \right)^2 \right] I(j \leq t)$$

The GKY kernel estimator: properties

- Consistency, need to assume
  - Persistence in parameters
  - Boundedness (up to bounded random walks)
- Computationally convenient:
  1. you can have larger n than in Bayesian TVP-VARs
  2. you can have larger p than in Bayesian TVP-VARs (monthly variables)

BUT HOW LARGE can n be?

- As n increases the GKY estimator will overfit
  - This is the main motivation of the paper
  - Can you regularise/penalize the GKY estimator?

Penalizing the GKY estimator *via* stochastic constraints

Basic idea: mix the GKY estimator with stochatic constraints, see Theil and Golbderger (1961)

$$\left(\underset{n\times 1}{y_t} \right) = \left(\underset{n\times nk}{I_n \otimes x_t'} \right) \underset{nk\times 1}{\beta_t} + \left(\underset{n\times 1}{u_t} \right)$$

Penalizing the GKY estimator *via* stochastic constraints

Basic idea: mix the GKY estimator with stochatic constraints, see Theil and Golbderger (1961)

$$\begin{pmatrix} \underset{n\times 1}{y_t} \\ \underset{nk\times 1}{\sqrt{\lambda} r} \end{pmatrix} = \begin{pmatrix} \underset{n\times nk}{I_n \otimes x_t'} \\ \sqrt{\lambda} \underset{nk\times nk}{R} \end{pmatrix} \underset{nk\times 1}{\beta_t} + \begin{pmatrix} \underset{n\times 1}{u_t} \\ \underset{nk\times 1}{u_t^r} \end{pmatrix}$$

Penalizing the GKY estimator *via* stochastic constraints

Basic idea: mix the GKY estimator with stochatic constraints, see Theil and Golbderger (1961)

$$\begin{pmatrix} \underset{n\times 1}{y_t} \\ \underset{nk\times 1}{\sqrt{\lambda} r} \end{pmatrix} = \begin{pmatrix} \underset{n\times nk}{I_n \otimes x_t'} \\ \sqrt{\lambda} \underset{nk\times nk}{R} \end{pmatrix} \underset{nk\times 1}{\beta_t} + \begin{pmatrix} \underset{n\times 1}{u_t} \\ \underset{nk\times 1}{u_t^r} \end{pmatrix}$$

- A number of estimators can be derived from a similar setup
  - Ridge estimator
  - Unbiased Ridge estimator
  - Dummy implementation of priors in Bayesian VARs
- Assumptions on u_t and u_t^r
  - u_t is a martingale difference sequence
  - $var(u_t) = \Sigma_n$, $var(u_t^r) = I_k \otimes \Sigma_n$

Penalizing the GKY estimator *via* stochastic constraints

$$\begin{pmatrix} \underset{n\times 1}{y_t} \\ \underset{nk\times 1}{\sqrt{\lambda} r} \end{pmatrix} = \begin{pmatrix} \underset{n\times nk}{I_n \otimes x_t'} \\ \sqrt{\lambda} \underset{nk\times nk}{R} \end{pmatrix} \underset{nk\times 1}{\beta_t} + \begin{pmatrix} \underset{n\times 1}{u_t} \\ \underset{nk\times 1}{u_t^r} \end{pmatrix}$$

Now simply apply the GKY estimator to the augmented model

$$\widehat{\beta}_t = \left[\underbrace{\left(I_n \otimes \sum_{j=1}^{T} w_{j,t}^H x_j x_j' \right)}_{nk\times nk} + \lambda R' R \right]^{-1} \left[\sum_{j=1}^{T} w_{j,t}^H \left(I_n \otimes x_j \right) y_j + \lambda R' r \right]$$

Penalizing the GKY estimator *via* stochastic constraints

$$\begin{pmatrix} \underset{n\times 1}{y_t} \\ \underset{nk\times 1}{\sqrt{\lambda} r} \end{pmatrix} = \begin{pmatrix} \underset{n\times nk}{I_n \otimes x_t'} \\ \sqrt{\lambda} \underset{nk\times nk}{R} \end{pmatrix} \underset{nk\times 1}{\beta_t} + \begin{pmatrix} \underset{n\times 1}{u_t} \\ \underset{nk\times 1}{u_t^r} \end{pmatrix}$$

Now simply apply the GKY estimator to the augmented model

$$\widehat{\beta}_t = \left[\underbrace{\left(I_n \otimes \sum_{j=1}^{T} w_{j,t}^H x_j x_j' \right)}_{nk\times nk} + \lambda R' R \right]^{-1} \left[\sum_{j=1}^{T} w_{j,t}^H \left(I_n \otimes x_j \right) y_j + \lambda R' r \right]$$

- $\lambda = 0$ $\qquad\qquad \widehat{\beta}_t = \widehat{\beta}_{t,GKY}$

Penalizing the GKY estimator *via* stochastic constraints

$$\begin{pmatrix} \underset{n\times 1}{y_t} \\ \underset{nk\times 1}{\sqrt{\lambda} r} \end{pmatrix} = \begin{pmatrix} \underset{n\times nk}{I_n \otimes x_t'} \\ \sqrt{\lambda} \underset{nk\times nk}{R} \end{pmatrix} \underset{nk\times 1}{\beta_t} + \begin{pmatrix} \underset{n\times 1}{u_t} \\ \underset{nk\times 1}{u_t^r} \end{pmatrix}$$

Now simply apply the GKY estimator to the augmented model

$$\widehat{\beta}_t = \left[\underbrace{\left(I_n \otimes \sum_{j=1}^{T} w_{j,t}^H x_j x_j' \right)}_{nk\times nk} + \lambda R' R \right]^{-1} \left[\sum_{j=1}^{T} w_{j,t}^H \left(I_n \otimes x_j \right) y_j + \lambda R' r \right]$$

- $\lambda = 0$ $\qquad \widehat{\beta}_t = \widehat{\beta}_{t,GKY}$
- $\lambda \to \infty$ $\qquad \widehat{\beta}_t \to \beta_C = (R'R)^{-1}(R'r)$

Penalizing the GKY estimator *via* stochastic constraints

$$\begin{pmatrix} \underset{n\times 1}{y_t} \\ \underset{nk\times 1}{\sqrt{\lambda} r} \end{pmatrix} = \begin{pmatrix} \underset{n\times nk}{I_n \otimes x_t'} \\ \sqrt{\lambda} \underset{nk\times nk}{R} \end{pmatrix} \underset{nk\times 1}{\beta_t} + \begin{pmatrix} \underset{n\times 1}{u_t} \\ \underset{nk\times 1}{u_t^r} \end{pmatrix}$$

Now simply apply the GKY estimator to the augmented model

$$\widehat{\beta}_t = \left[\underbrace{\left(I_n \otimes \sum_{j=1}^{T} w_{j,t}^{H} x_j x_j' \right)}_{nk\times nk} + \lambda R' R \right]^{-1} \left[\sum_{j=1}^{T} w_{j,t}^{H} \left(I_n \otimes x_j \right) y_j + \lambda R' r \right]$$

- $\lambda = 0$ $\qquad \widehat{\beta}_t = \widehat{\beta}_{t,GKY}$
- $\lambda \to \infty$ $\qquad \widehat{\beta}_t \to \beta_C = (R'R)^{-1}(R'r)$
- The estimator depends on two constants, λ and H

Focus on cases where R has a kronecker structure

If the constraints are such that R has a kronecker structure the model can be cast in matrix form.
Use the following definitions:

- $\underset{nk\times nk}{R} = (I_n \otimes \underset{k\times k}{\overline{R}})$, $\underset{kn\times 1}{u_t^r} = vec(\underset{k\times n}{\overline{u}_t^r})$, $\underset{kn\times 1}{r} = vec(\underset{k\times n}{\overline{r}})$

$$\begin{pmatrix} \underset{1\times n}{y_t'} \\ \underset{k\times n}{\sqrt{\lambda}\overline{r}} \end{pmatrix} = \begin{pmatrix} \underset{1\times k}{x_t'} \\ \underset{k\times k}{\sqrt{\lambda}\overline{R}} \end{pmatrix} \underset{k\times n}{\Theta_t} + \begin{pmatrix} \underset{1\times n}{u_t'} \\ \underset{k\times n}{\overline{u}_t^r} \end{pmatrix}$$

which we can re-write as:

$$\underset{(k+1)\times n}{y_t^*} = x_t'^*\Theta_t + u_t^*,$$

$$\underset{k\times n}{\widehat{\Theta}_t} = \underbrace{\left(\sum_{j=1}^{T} w_{j,t} x_j^* x_j^{*\prime}\right)}_{k\times k}^{-1} \left(\sum_{j=1}^{T} w_{j,t} x_j^* y_j^*\right),$$

The penalized GKY estimator as a linear combination of two estimators

- The estimator can be written as

$$\widehat{\beta}_t = A\widehat{\beta}_{t,GKY} + (I - A)\beta_C$$

- Hence it is a linear combination of:

The penalized GKY estimator as a linear combination of two estimators

- The estimator can be written as

$$\widehat{\beta}_t = A\widehat{\beta}_{t,GKY} + (I - A)\beta_C$$

- Hence it is a linear combination of:
- the unbiased (time varying) GKY estimator

The penalized GKY estimator as a linear combination of two estimators

- The estimator can be written as

$$\widehat{\beta}_t = A\widehat{\beta}_{t,GKY} + (I - A)\beta_C$$

- Hence it is a linear combination of:
- the unbiased (time varying) GKY estimator
- a time invariant structure imposed by the constraints

The penalized GKY estimator as a linear combination of two estimators

- The estimator can be written as

$$\widehat{\beta}_t = A\widehat{\beta}_{t,GKY} + (I - A)\beta_C$$

- Hence it is a linear combination of:
- the unbiased (time varying) GKY estimator
- a time invariant structure imposed by the constraints
- This helps in understanding the bias/variance trade off

Bias/variance trade off in the penalized GKY estimator

Theorem (asymptotic distribution/bias)

Let

- u_t *be a martingale difference sequence with finite fourth moments*
- *the coefficients evolve slowly*

$$\sup_{j \leq h} \left\| \beta_t - \beta_{t+j} \right\| = O_p\left(\frac{h}{t}\right).$$

- $H = o(T^{1/2})$

Then

$$\left(\Gamma_{w,t}^{*-1} \Gamma_{ww,t}^{**} \Gamma_{w,t}^{*-1} \otimes \Sigma_n\right)^{-\frac{1}{2}} \sqrt{H}\left(\widehat{\beta}_t - \beta_t - \beta_t^B\right) \to^d N(0, I)$$

$$\beta_t^B = p\lim S_w^{-1} \lambda \overline{R}'(r - R\beta_t)$$

Bias/variance trade off in the penalized GKY estimator

Theorem (variance)

Let

$$\widehat{\beta}_t = [E+F]^{-1}\left[E\widehat{\beta}_{t,GKY} + \lambda R' r\right]$$

Then

$$var(\widehat{\beta}_t) = [E+F]^{-1}\left[E var(\widehat{\beta}_{t,GKY})E\right][E+F]^{-1}$$

and for any vector q

$$q'\left(var(\widehat{\beta}_{t,GKY}) - var(\widehat{\beta}_t)\right)q \geq 0$$

is a positive semi-definite matrix: i.e. the penalty induces a variance reduction

Time varying volatilities

If the VAR residuals have time varying covariance matrix, follow the two step procedure by Giraitis et al. (2012)

$$\hat{\Psi}_t = \sum_{j=1}^{T} w_{j,t}(H_\Psi) u_t u_t'$$

and use a GLS correction to obtain

$$\beta_t = \left[\sum_{j=1}^{T} w_{j,t} \left(\hat{\Psi}_j^{-1} \otimes x_j' x_j\right) + R' \widetilde{\Psi_t}^{-1} R\right]^{-1} \left[\sum_{j=1}^{T} w_{j,t} vec\left(x_j' y_j' \hat{\Psi}_j^{-1}\right) + R' \widetilde{\Psi_t}^{-1} r\right]$$

- Computation slows down: now have to invert *nk* dimensional matrices
- In an empirical application we do not find any advantage from applying this GLS correction

Ridge/Litterman type estimators

Ridge penalty

$$r = \begin{pmatrix} 0_{k\times n} \end{pmatrix} \qquad \overline{R} = I_k$$

Litterman penalty

$$\overline{r} = \begin{pmatrix} \mathrm{diag}(\delta_1\sigma_1, \delta_2\sigma_2, \delta_3\sigma_3, ..., \delta_n\sigma_n) \\ 0_{n(p-1)+1\times n} \end{pmatrix} \qquad \overline{R} = \begin{pmatrix} \overline{\Sigma} & 0 \\ 0 & \sigma_c^2 \end{pmatrix}, \text{ where}$$

$$\overline{\Sigma} = diag(1, 2, 3, ..., p) \otimes diag(\sigma_1, \sigma_2, ..., \sigma_n)$$

Selecting λ and H: Method 1

- Adapt the method by Banbura, Giannone, and Reichlin (2010)
- Intuition: penalize the overfitting of the large model up to the point where you achieve the same fit as a small system

The loss function to be minimized is:

$$L_{fit}(\lambda, H) = \left| \sum_{i=1}^{n_1} \frac{rss_n^i(\lambda, H)}{rss_{RW}^i} - \sum_i \frac{rss_{n1}^i}{rss_{RW}^i} \right|$$

where n1 is a subset of variables of interest

Selecting λ and H: Method 2

- The second method aims at penalizing poor average forecasting performance over recent observations
- At each step t in the forecast exercise consider a relatively short window of recent data $t-L-h, t-L-h+1, \ldots, t-1-h$, then

The loss function to be minimized is:

$$L_{mse}(\lambda, H) = \sum_{i=1}^{n_1} \frac{mse^{i}_{h,L}(\lambda, H)}{var(y_{t,i})}$$

Practical implementation

- Estimator is easy to compute
- Feasible solution is represented by a grid search approach
- We use a wide (38 elements) grid for (the reciprocal) of λ, $\varphi = 1/\lambda$

$$\varphi_{grid} = 10^{-10}, 10^{-5}, 10^{-4}, 10^{-3}, ..., 1$$

- s for the kernel $w_{j,t}$, we use a 6 points grid for H:

$$H_{grid} = 0.5, 0.6, 0.7, 0.8, 0.9, 1,$$

- Yet the best results are obtained by pooling, either with equal weights or with weights based on the above criteria

Finite sample properties

DGP-1: random walk coefficients [0.85 1] + random walk volatilities

$$Y_t = \Psi_t Y_{t-1} + \varepsilon_t$$
$$\Psi_t = \Psi_{t-1} + \eta_t$$

DGP-2: occasionally breaking coefficients [0 1] + random walk volatilitiess

$$Y_t = \Psi_t Y_{t-1} + \varepsilon_t$$
$$\Psi_t = (1 - I(\tau))\Psi_{t-1} + I(\tau)\Psi_{t-1}\eta_t$$

- τ governs the probability of coefficients changing (on average once every 10 years)

DGP-3: sine shaped coefficients [-1 1] + random walk volatilities

$$Y_t = \Psi_t Y_{t-1} + \varepsilon_t$$
$$\Psi_t = sin(10\pi t/T) + \eta_t$$

In all DGPs we assume random walk stochastic volatilities

$$\begin{aligned} \varepsilon_{it} &= u_{it} \exp(\lambda_{it}) \\ \lambda_{it} &= \lambda_{it-1} + \eta_{it} \end{aligned}$$

where $u_{it} \backsim N(0,1)$ and $\eta_{it} \backsim N(0,\sigma_\eta)$. We calibrate $\sigma_\eta = 0.01$

Comparison with parametric estimator

- Alternative approach: (TVP-VAR) by Koop and Korobilis (2013)
- Full parametric specification

$$\begin{aligned} y_t &= Z_t \beta_t + \varepsilon_t,\ \varepsilon_t \sim N(0, \Sigma_t) \\ \beta_{t+1} &= \beta_t + u_{t+1},\ u_{t+1} \sim N(0, Q_{t+1}) \end{aligned}$$

- To obtain the β_t need estimate of Q_t and Σ_t from Kalman filter
- Bayesian estimation with MCMC unfeasible for $n > 4, 5$
- Hint: what binds is the size of the state vector β_t, which is $n(np+1)$
- Koop and Korobilis (2013) propose a solution following Ljung (1992) and Sargent (1999)

Forgetting factors

Simplification 1

$$Q_t = \left(\frac{1-\gamma}{\gamma}\right) P_{t-1/t-1}$$

- $P_{t-1/t-1}$ comes from the Kalman filter: $cov(\beta_{t-1}/I_{t-1})$
- Time variation Q_t is a fraction of the uncertainty on β_t

Simplification 2

$$\widehat{\Sigma}_t = \kappa\widehat{\Sigma}_{t-1} + (1-\kappa)v_t v_t'$$

- v_t is the prediction error from the Kalman filter

Hence

- The Kalman filter gives β_t conditional on Q_t and Σ_t
- But Q_t and Σ_t are themselves function of the Kalman filter output
- Everything driven by two constants γ and κ + Initial condition

Main differences with the parametric approach

1. Driftless random walk assumption for computational convenience. If the true DGPs is very different it could result in poor performance.
2. Curse of dimensionality: for 20 variables and 4 lags (KK application with quarterly data) β_t contains 1620 elements. Monthly models untractable
3. The *prior* only on the initial condition β_1 dies out relatively quickly. Also, the longer the sample size, the lower the effect of the prior on the parameter estimates. In our estimator the stochastic constraints are effective at each point in time.

But in practice?

Table: 1 step ahead, relative RMSEs

T	Parametric	Non parametric	
		Inv. RMSE	Equal weights
DGP 1 (Random walk coefficients)			
n=7			
100	1	1.004	1.005
150	1	0.999	1.000
200	1	0.997	0.997
n=15			
100	1	1.021	1.024
150	1	1.012	1.013
200	1	1.006	1.007
DGP 2 (Occasionally breaking coefficients)			
n=7			
100	1	0.96	0.96
150	1	0.96	0.96
200	1	0.96	0.96
n=15			
100	1	0.96	0.96
150	1	0.95	0.95
200	1	0.94	0.94
DGP 3 (Sine function coefficients)			
n=7			
100	1	0.95	0.96
150	1	0.96	0.98
200	1	0.97	0.99
n=15			
100	1	0.87	0.88
150	1	0.86	0.87
200	1	0.85	0.86

Empirical application 1: forecasting

- Pseudo real time forecasting exercise: monthly data from 1960 to 2013
- Forecast CPI/Employment/Fed Fund Rates 1 to 24 steps ahead
- Experiment with datasets of different sizes: n=20, 78
- We organize the forecast exercise around three questions
- Q.1 Does time variation actually improve forecast accuracy?

Empirical application 1: forecasting

- Pseudo real time forecasting exercise: monthly data from 1960 to 2013
- Forecast CPI/Employment/Fed Fund Rates 1 to 24 steps ahead
- Experiment with datasets of different sizes: n=20, 78
- We organize the forecast exercise around three questions
- Q.1 Does time variation actually improve forecast accuracy?

YES

Empirical application 1: forecasting

- Pseudo real time forecasting exercise: monthly data from 1960 to 2013
- Forecast CPI/Employment/Fed Fund Rates 1 to 24 steps ahead
- Experiment with datasets of different sizes: n=20, 78
- We organize the forecast exercise around three questions
- Q.1 Does time variation actually improve forecast accuracy?

YES

- Q.2 Can the performance of *medium-sized* TVP-VARs be approximated by that of *large* VARs with *constant coefficients* (Stock and Watson, 2012, Aastveit, Carriero, Clark and Marcellino, 2014)

Empirical application 1: forecasting

- Pseudo real time forecasting exercise: monthly data from 1960 to 2013
- Forecast CPI/Employment/Fed Fund Rates 1 to 24 steps ahead
- Experiment with datasets of different sizes: n=20, 78
- We organize the forecast exercise around three questions
- Q.1 Does time variation actually improve forecast accuracy?

  YES

- Q.2 Can the performance of *medium-sized* TVP-VARs be approximated by that of *large* VARs with *constant coefficients* (Stock and Watson, 2012, Aastveit, Carriero, Clark and Marcellino, 2014)

  NOT (at long horizons)

Empirical application 1: forecasting

- Pseudo real time forecasting exercise: monthly data from 1960 to 2013
- Forecast CPI/Employment/Fed Fund Rates 1 to 24 steps ahead
- Experiment with datasets of different sizes: n=20, 78
- We organize the forecast exercise around three questions
- Q.1 Does time variation actually improve forecast accuracy?

  YES

- Q.2 Can the performance of *medium-sized* TVP-VARs be approximated by that of *large* VARs with *constant coefficients* (Stock and Watson, 2012, Aastveit, Carriero, Clark and Marcellino, 2014)

  NOT (at long horizons)

- Q.3 Does it pay off to go beyond a medium size system, i.e. does going from a 20 to a 78 TVP-VAR improve forecast accuracy?

Empirical application 1: forecasting

- Pseudo real time forecasting exercise: monthly data from 1960 to 2013
- Forecast CPI/Employment/Fed Fund Rates 1 to 24 steps ahead
- Experiment with datasets of different sizes: n=20, 78
- We organize the forecast exercise around three questions
- Q.1 Does time variation actually improve forecast accuracy?

  YES

- Q.2 Can the performance of *medium-sized* TVP-VARs be approximated by that of *large* VARs with *constant coefficients* (Stock and Watson, 2012, Aastveit, Carriero, Clark and Marcellino, 2014)

  NOT (at long horizons)

- Q.3 Does it pay off to go beyond a medium size system, i.e. does going from a 20 to a 78 TVP-VAR improve forecast accuracy?

  NOT

Q.1 Does time variation improve forecast accuracy?

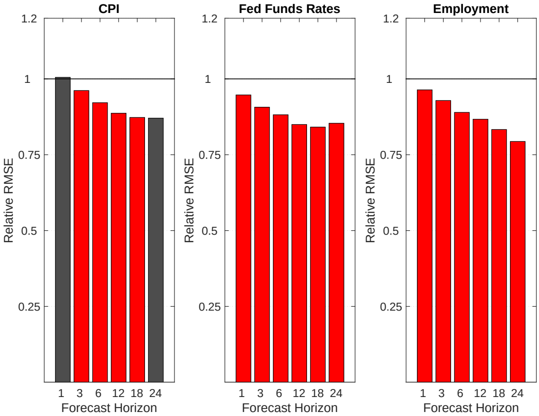


Figure: Root Mean Square Forecast Errors: combined TVP-VARs versus constant coefficients BVAR (20 variables VARs)

Q.1 Does time variation improve forecast accuracy?

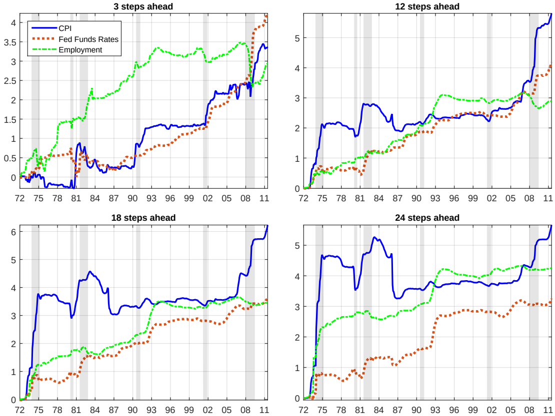


Figure: Cumulative sum of squared forecast error differentials: combined TVP-VARs versus constant coefficients BVAR (20 variables VARs)

Q.2 is time variation due to omitted variables?

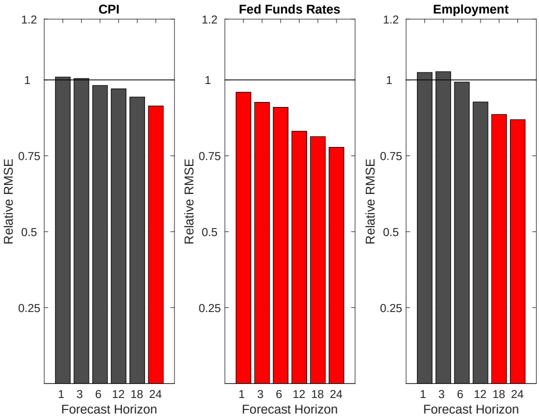


Figure: Root Mean Square Forecast Errors: 20 variables combined TVP-VARs versus 78 variables constant coefficients BVAR

Q.2 is time variation due to omitted variables?

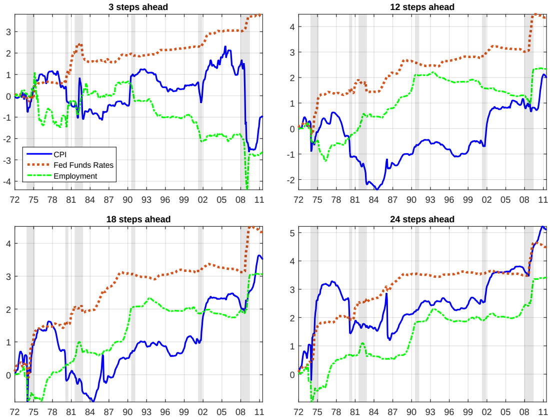


Figure: Cumulative sum of squared forecast error differentials: 20 variables combined TVP-VARs versus 78 variables constant coefficients BVAR

Q.3 do we need more than n=20?

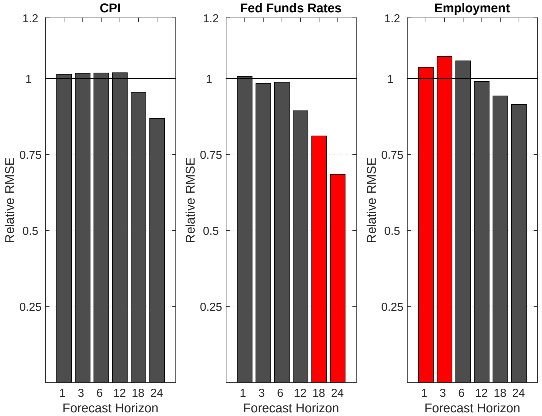


Figure: Root Mean Square Forecast Errors: 20 variables combined TVP-VARs versus 78 variables combined TVP-VARs

Q.3 do we need more than n=20?

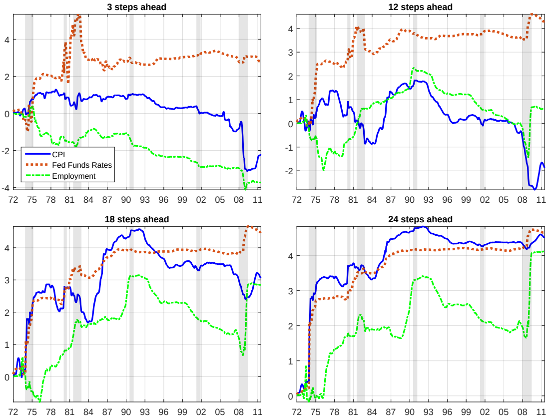


Figure: Cumulative sum of squared forecast error differentials: 20 variables combined TVP-VARs versus 78 variables combined TVP-VARs

Comparison with the parametric estimator

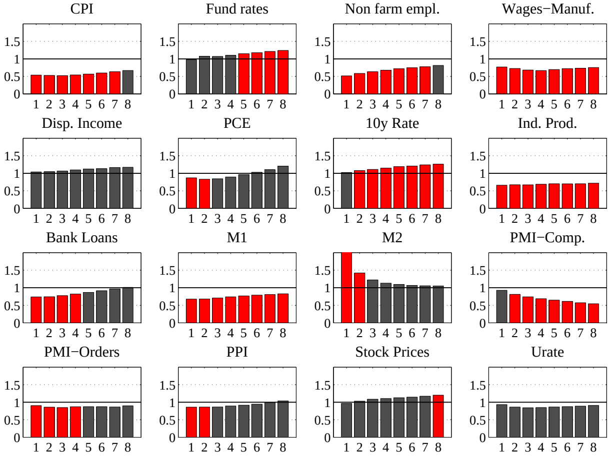


Figure: Forecast accuracy, nonparametric and parametric estimators

Structural analysis with large models

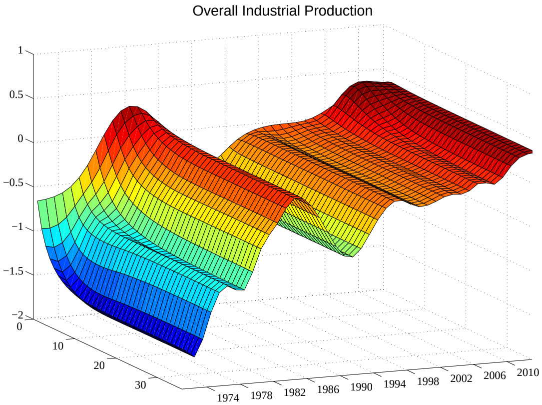
- Add granularity to the evidence of instability in the oil price/macroeconomy relationship, Edelstein and Kilian (2007), Blanchard and Gali (2009), Baumeister and Peersman (2013)
- Some authors have emphasized improvements in car efficiency as a crucial factor for lower impact of oil price shocks on output and decreasing elasticity of oil demand mpg
- We assess this channel by looking at the time varying responses of sectorial IP to an oil price shock
- Strategy: augment the baseline 20 variables VAR with 8 industrial production series split by product destination including durable consumption (mostly vehicles)
- Time varying volatilities modeled in two steps like in Giraitis et al. (2012)

Industrial production Indexes by market group

Market group	Acronym	Weight
Industrial Production Index	INDPRO	100
Industrial Production: Business Equipment	IPBUSEQ	9.18
Industrial Production: Consumer Goods	IPCONGD	27.2
Industrial Production: Durable Consumer Goods	*IPDCONGD*	*5.59*
Industrial Production: Nondurable Consumer Goods	*IPNCONGD*	*21.62*
Industrial Production: Final Products (Market Group)	IPFINAL	16.58
Industrial Production: Materials	IPMAT	47.03
Industrial Production: Durable Materials	*IPDMAT*	*17.34*
Industrial Production: Nondurable Materials	*IPNMAT*	*11.44*

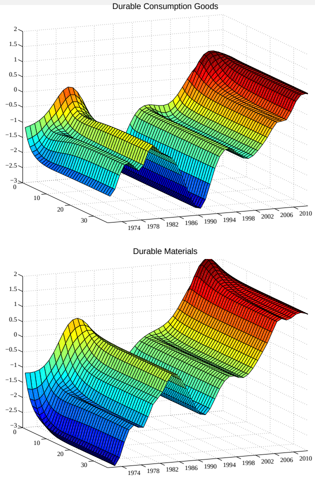
Table: Industrial production indexes by market group

Overall industrial output



Response of Industrial production (overall index) to a 1% real oil price innovation

Bus. equipment & Durable Materials as relevant as Durable consumption



Response of Industrial production (selected market groups) to a 1% real oil price innovation

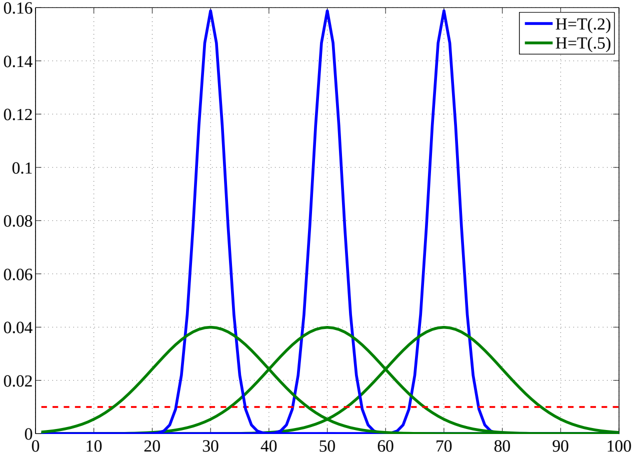
Conclusions

- We propose a nonparametric estimator for large TVP-VAR
- The estimator is a penalized version of the GKY estimator
- Like every penalized estimator it offers a bias/variance trade-off
- Selection of the penalty parameter and of the bandwidth via a number of cross-validation strategies, but pooling works very well in practice
- Forecast accuracy: it outperforms constant parameter benchmarks
- Compared to a parametric estimator:
    1. It overcomes limitations in terms of size
    2. Compares favourably in terms of actual forecast accuracy
    3. In Monte Carlo exercises it proves robust to heteroschedastic errors and different specifications for the VAR coefficients
- It proves useful for structural analysis

THANKS!

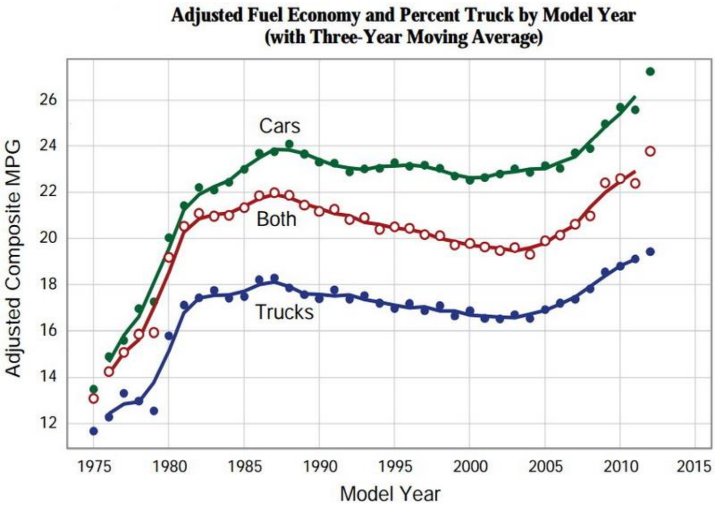
Example: gaussian kernel function

back



Average Mileage per Gallon (MPG)

back



Share of vehicles by Mileage per Gallon (MPG)

back