ESCoE Research Seminar

Estimating High Frequency Regional Output Growth Using Mixed Frequency Methods

Presented by James Mitchell, Warwick Business School

12 June 2018

Estimating High Frequency Regional Output Growth Using Mixed Frequency Methods

Gary Koop (Strathclyde, ESCoE), Stuart G. McIntyre (Strathclyde, ESCoE), James Mitchell (Warwick, ESCoE) & Aubrey Poon (Strathclyde, ESCoE)

ESCoE Seminar, NIESR

12 June 2018

Background: Frequency and Timeliness

1. Regional UK nominal GVA data are currently available at annual frequency
2. Having regional data at higher frequency (e.g. quarterly) desirable
3. Regional UK GVA data are currently released with long delay (right now only have 2016)
4. Desirable to have quicker estimates
5. Other data are available more frequently and in a more timely fashion
6. GVA for UK as a whole are available quarterly and shorter release delay (right now have 2018Q1)
7. Can we use these higher frequency and more timely data to produce estimates (historical, present and future) of regional GVA growth?

Mixed Frequency Econometrics

1. Growing literature on building econometric models which combine data of different frequencies
2. Schorfheide and Song (2015). Real-time forecasting with a mixed-frequency VAR. Journal of Business Economic Statistics
3. Ghysels (2016). Macroeconomics and the reality of mixed frequency data. Journal of Econometrics
4. Brave, Butters and Justiniano (2016). Forecasting economic activity with mixed frequency Bayesian VARs. Federal Reserve Bank of Chicago Working Paper
5. McCracken, Owyang and Sekhposyan (2016). Real-time forecasting with a large, mixed frequency Bayesian VAR, Federal Reserve Bank of St. Louis Working Paper

Mixed Frequency Econometrics

1. Intuition: Estimate econometric model using high and low frequency variables
2. Use these estimated relationships to "fill in" values for the low frequency variables
3. We build on approach of Schorfheide and Song (2015)
4. Advantage 1: Can predict UK regional GVA at time t+h given data available at time t
5. Useful for forecasting/nowcasting providing timely flash estimates
6. Advantage 2: Can provide historical estimates (estimate time t given all data up to time T)
7. Competing approach is stacked VAR: Has Advantage 1 but not Advantage 2

The Mixed Frequency Vector Autoregressive Model: MF-VAR

1. Schorfheide and Song (2015) is a state space model involving a VAR
2. First define their model
3. Show why we thought it would be too tough to estimate in our UK regional application
4. Then show new features we added to obtain estimable model:
5. Incorporate the cross-sectional restriction that weighted average of regional GVA growth = UK GVA growth
6. Machine learning methods for Big Data (Dirichlet Laplace hierarchical prior)

MF-VAR: Notation

1. $t = 1, .., T$ runs at the *quarterly* frequency.
2. $r = 1, .., R$ denotes the R regions in the UK.
3. y_t^{UK} is the quarterly change in GVA in the UK. We observe this.
4. $y_t^{r,A}$ is annual GVA growth in region r. It is observed, but only in quarter 4 of each year.
5. y_t^r is the quarterly change in GVA in region r. It is never observed.
6. y_t is a $n = R + 1$ vector containing quarterly growth rates. The regional ones are not observed.

MF-VAR-SV: Definition

1. The MF-VAR is a VAR using y_t as the dependent variables:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + ... + \Phi_p y_{t-p} + u_t \tag{1}$$

2. u_t is i.i.d. $N(0, \Sigma_t)$
3. Note: it is Σ_t not Σ
4. We allow for multivariate stochastic volatility (SV)
5. Empirically important for density forecasts
6. Thus we have MF-VAR-SV

MF-VAR-SV: Definition

1. This VAR contains lots of missing observations: we do not observe quarterly regional growth rates
2. We do observe (every fourth quarter) the annual regional growth rate
3. Relationship between observed annual growth rate to unobserved quarterly growth rate is approximately:

$$y_t^{r,A} = \frac{1}{4}y_t^r + \frac{1}{2}y_{t-1}^r + \frac{3}{4}y_{t-2}^r + y_{t-3}^r + \frac{3}{4}y_{t-4}^r + \frac{1}{2}y_{t-5}^r + \frac{1}{4}y_{t-6}^r \quad (2)$$

4. Call this the inter-temporal restriction
5. We also observe y_t^{UK}

Estimating the MF-VAR-SV

1. Estimation of the MF-VAR-SV proceeds by treating it as a state space model
2. The VAR in equation (1) is state equation
3. Unknown quarterly regional growth rates = states
4. Equation (2) is what we observe = measurement equation
5. Standard statistical methods exist for estimating such models
6. MF-VAR-SV also allows for other patterns, missing observations (e.g. due to release delays)
7. This produces estimates and forecasts of quarterly regional GVA growth
8. Not only point estimates, but also posterior and predictive densities

Why Do We not Use this Version of the Model?

1. Schorfheide and Song (and others) typically use many high frequency variables and few low frequency ones
2. E.g. 10 monthly macroeconomic variables and quarterly GDP growth
3. Our frequency mis-match is opposite: 12 annual variables and 1 quarterly variable
4. We have relatively short sample size
5. Annual regional growth rates from 1967 to 2016
6. We are trying to provide estimates of 200 regional growth rates for each of 12 regions
7. Lots of VAR parameters to estimate: with p=4 lags, almost 700 coefficients to estimate
8. And Σ_t has almost 100 parameters to estimate
9. And each period has a different Σ_t
10. Trying to estimate too much with too little data
11. In theory, MF-VAR-SV is estimable, but in practice?

What We Add: Machine Learning Methods

1. With Big Data often have similar problems
2. E.g. regression with many more predictors than observations
3. Machine learning methods very popular in such cases
4. We take a popular machine learning method: using Dirichlet-Laplace hierarchical prior
5. Automatically sorts through huge numbers of parameters and sets to zero irrelevant ones
6. Big gain in parsimony

What We Add: Cross-Sectional Restriction

1. We can draw on another source of information based on the fact that UK GVA is the sum of the regions
2. Call this cross-sectional restriction
3. How to incorporate this in the econometric model?
4. Doran, (1992). Constraining Kalman filter and smoothing estimates to satisfy time-varying restrictions. The Review of Economics and Statistics
5. This is exactly what we need to do
6. Allow for the cross-sectional restriction to hold stochastically

Econometric Methods: Summary

1. Our goal is to produce estimates of regional GVA growth y_t^r
2. To do so, our MF-VAR-SV uses information on:
3. Annual regional growth rate $y_t^{r,A}$
4. Quarterly UK growth rate y_t^{UK}
5. The cross-sectional restriction
6. The Dirichlet-Laplace prior
7. It is state space model + machine learning prior estimated using Markov Chain Monte Carlo methods

Data

1. Annual nominal and real GVA growth data from 1967 through 2016 (seasonally adjusted): **final vintage** and first release (for forecasting)
2. NUTS-1 regions
3. Data for NUTS-1 regions goes back to 1995; real GVA(B) data back to 1997
4. ONS nominal historical regional GVA data back to 1966; deflated (by UK deflator) for real database (but we allow for an error when imposing)
5. Statistical regions changed
6. Need to reconcile data prior to 1995
7. We also use extra UK quarterly variables in the MF-VAR
8. Inflation, interest rates, change in oil prices and exchange rates
9. Largest MF-VAR involves 17 variables

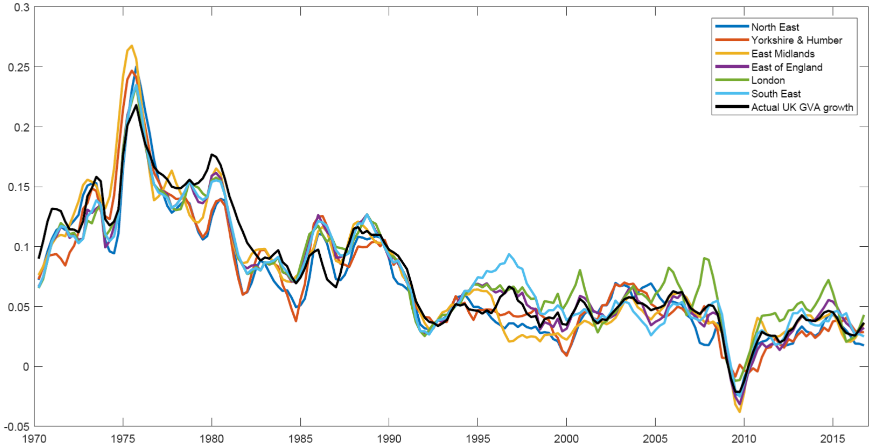
Model Choice

1. Model comparison exercise using marginal likelihoods
2. MF-VAR-SV using all 17 variables is best
3. Empirical results (with some exceptions) are for MF-VAR-SV
4. Strong evidence for SV
5. $p = 7$ lags

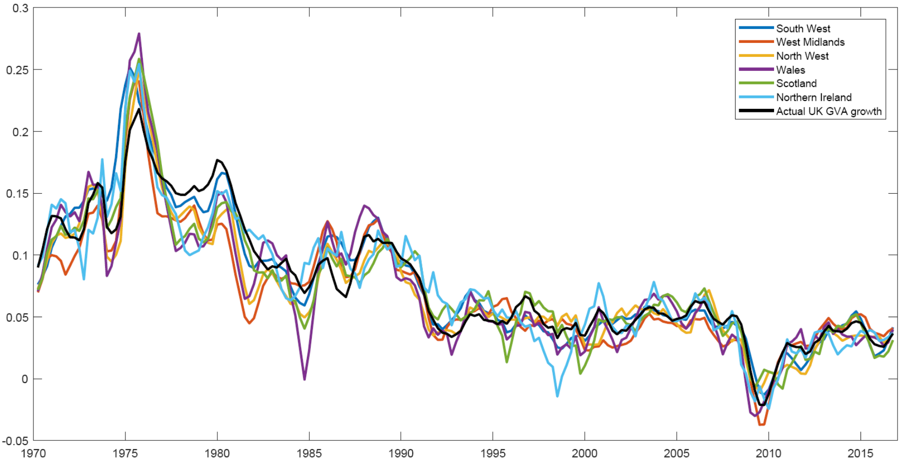
Historical Estimates of Regional GVA Growth

1. Next two figures take quarterly estimates y_t^r and convert to annual growth rates
2. Last quarter each year these estimates coincide with the actual annual regional growth rates (inter-temporal restriction)
3. Other quarters can differ
4. Within year variation in regional growth rates reflect UK GVA growth and other quarterly variables
5. These are point estimates, but credible intervals are narrow indicating precise estimation

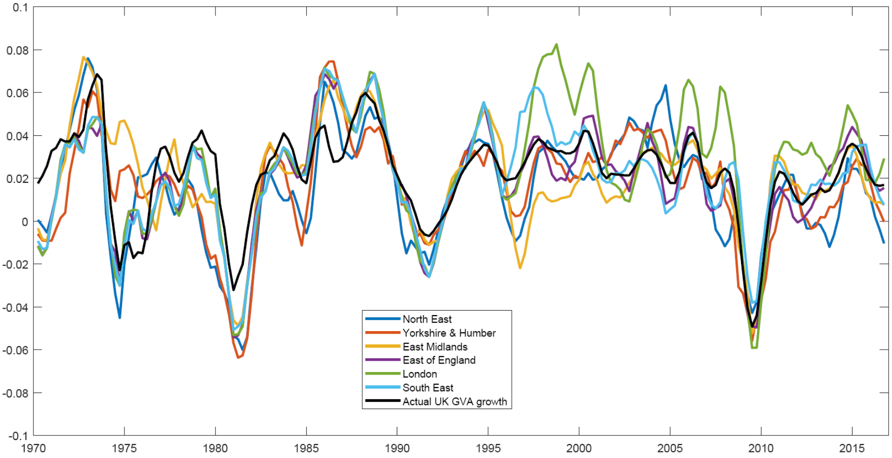
Nominal GVA Growth and Estimates for the UK Regions



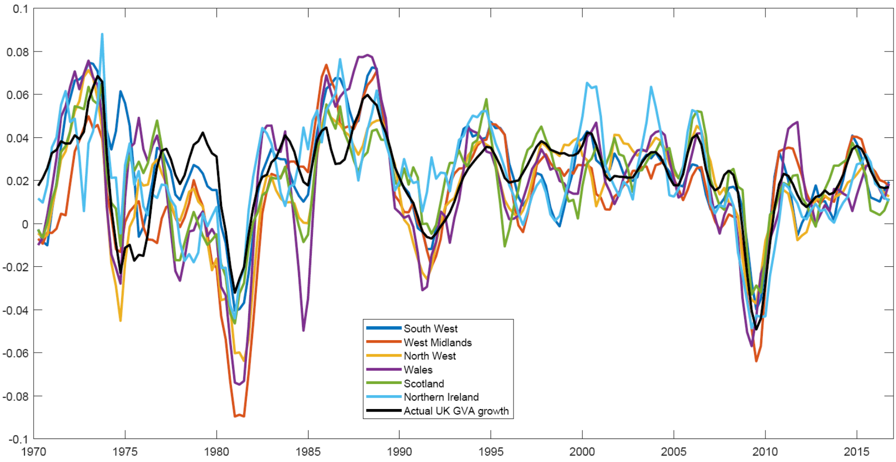
Nominal GVA Growth and Estimates for the UK Regions



Real GVA Growth and Estimates for the UK Regions



Real GVA Growth and Estimates for the UK Regions



Connectedness of UK Regions

1. Connectedness measure of Diebold and Yilmaz (2014, Journal of Econometrics)
2. Construct forecast error variance decompositions $d_{i,j}^h$ for $i, j = 1, .., n$ and $h = 1, .., H$
3. Proportion of the h-step ahead forecast error for region i which is accounted for by the errors in the equation for region j
4. We use generalised variance decomposition
5. Total connectedness FROM other regions to region i at horizon h is

$$\sum_{j \neq i} d_{i,j}^h$$

6. Measure of how information in other regions impacts the forecast error variance of region i
7. Call this "connectedness from" measure
8. "Connectedness to" is

$$\sum_{i \neq j} d_{i,j}^h$$

Connectedness of UK Regions

1. Our connectedness measures are at the quarterly frequency
2. $h = 1$ measure connectedness in terms of the one quarter ahead forecast error variances
3. These could not have been produced using VAR with annual data
4. Our connectedness measures are time varying; next table for 2016Q4 for MF-VAR-SV
5. Substantial variation across regions
6. Substantial differences between To and From
7. Even at very short ($h = 1$) horizon large amount of connectedness
8. At longer horizons ($h = 20$) connectendess measures rise

From and To Connectedness Measures in 2016Q4: Nominal GVA

	From h=1	To h=1	From h=20	To h=20
UK	34%	10%	81%	73%
CPI	28%	6%	77%	26%
Bank rate	24%	6%	77%	24%
exchange rate	30%	23%	79%	52%
oil	31%	84%	82%	147%
North East	30%	19%	89%	58%
Yorkshire and The Humber	24%	16%	90%	51%
East Midlands	42%	38%	89%	79%
East of England	40%	48%	85%	104%
London	31%	34%	84%	91%
South East	34%	27%	87%	71%
South West	31%	24%	87%	71%
West Midlands	27%	23%	86%	73%
North West	35%	39%	81%	121%
Wales	32%	53%	78%	141%
Scotland	28%	34%	81%	105%
Northern Ireland	26%	42%	78%	123%

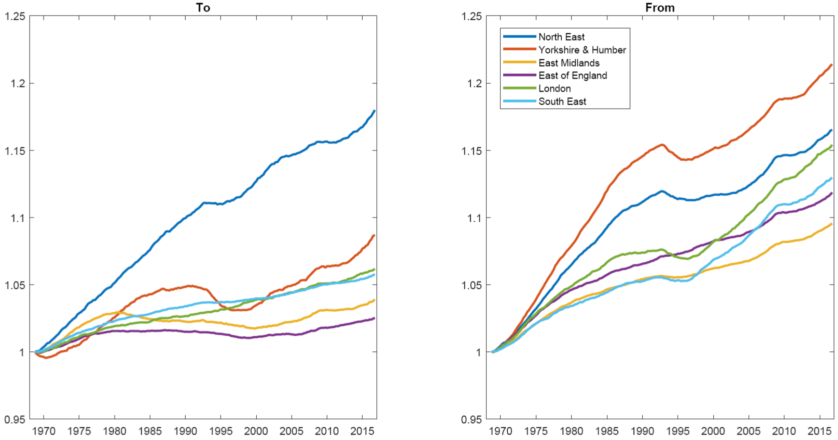
From and To Connectedness Measures in 2016Q4: Real GVA

	From	To	From	To
	h=1	h=1	h=20	h=20
UK	31%	8%	79%	67%
CPI	22%	6%	70%	25%
Bank rate	21%	6%	71%	28%
exchange rate	29%	24%	78%	56%
oil	29%	85%	81%	149%
North East	28%	18%	88%	61%
Yorkshire and The Humber	22%	16%	88%	54%
East Midlands	37%	31%	88%	72%
East of England	35%	41%	84%	100%
London	27%	30%	84%	83%
South East	31%	26%	87%	72%
South West	28%	23%	86%	68%
West Midlands	26%	20%	85%	70%
North West	32%	35%	80%	117%
Wales	30%	46%	77%	137%
Scotland	28%	32%	80%	106%
Northern Ireland	25%	39%	78%	118%

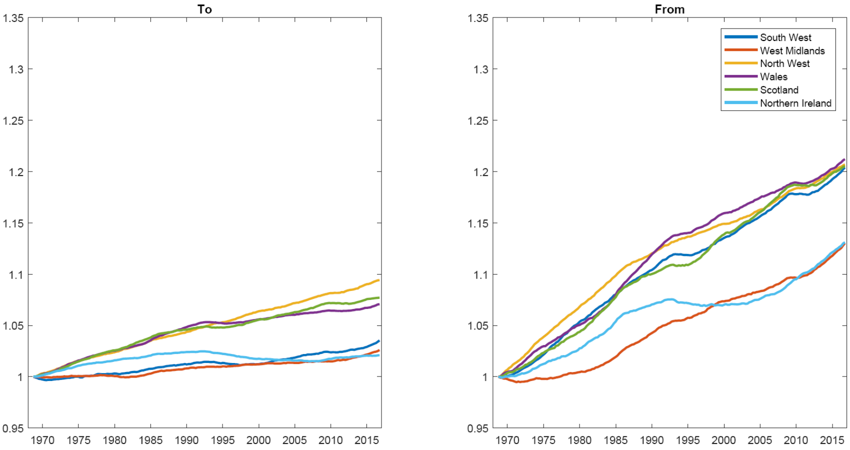
Connectedness of UK Regions

1. Preceding connectedness results for 2016Q4, but there is variation over time
2. To examine time variation make indices for each region (1970=100) for $h = 1$ (real GVA)
3. Increasing connectedness over time in both To and From measures
4. Regional connectedness grows even more strongly over time if we ignore the macro variables (especially for the North East and Yorks)
5. At h=20 From more constant over time and To has decreased over time (especially for the South West)

Change in Connectedness Over Time, h=1 (real GVA)



Change in Connectedness Over Time, h=1 (real GVA)



Conclusions and Extensions

1. MF-VAR-SV with Dirichlet-Laplace prior works well for producing quarterly regional estimates
2. Historical quarterly regional GVA growth estimates are useful in and of themselves
3. Also useful to examine (high frequency) connectedness between regions and macro variables
4. Current plan: Forecasting and (timely) nowcasting
5. Future plan: Regular quarterly production and publication of regional nowcasts (part of ESCoE output)