Forecasting Economic and Financial Time Series
Lecture Week 7: Big data and factor models

Ron Smith
MSc/PGCE Option, EMS Birkbeck

Spring 2021

Six questions for forecasters

1. **Decision environment and loss function. Why are we doing it?** A good forecast is one that leads to good decisions, judge this by some **Loss Function.**
2. **Forecast object. What are we forecasting?** A time series, such as sales, or an event such as a recession.
3. **Forecast Statement. How to we wish to state our forecast?** Point, interval or density forecast?
4. **Forecast Horizon. How far into the future do we wish to forecast?** This has implications for modelling strategy.
5. **Information set. What information is available at the time of forecasting?**, How much past history of the series? Other data? How big an estimation window?
6. **Method and complexity. Which method to use and how complex should the forecasting model be?** Parsimony (few estimated parameters) and shrinkage towards a prior may help.

Information Set: big data

- Lots of new sources of data, e.g. Google Trends. Night-lights data from satelites. AB testing. May help in forecasting.
- Bank of England using clearing/credit card data during the pandemic since most of the usual surveys not working
- Data Issues traditionally begin with V: Volume, Velocity (real time, stream not batch), Variety, Veracity, Value & Visualisation
- Computing constraints reduced but still there.
- Large number of pieces of data, e.g. very high frequency (nanosecond) data in finance.
- More explanatory variables (features) than observations
- Non stationary economic data complicates the cross-validation which is common in machine learning.

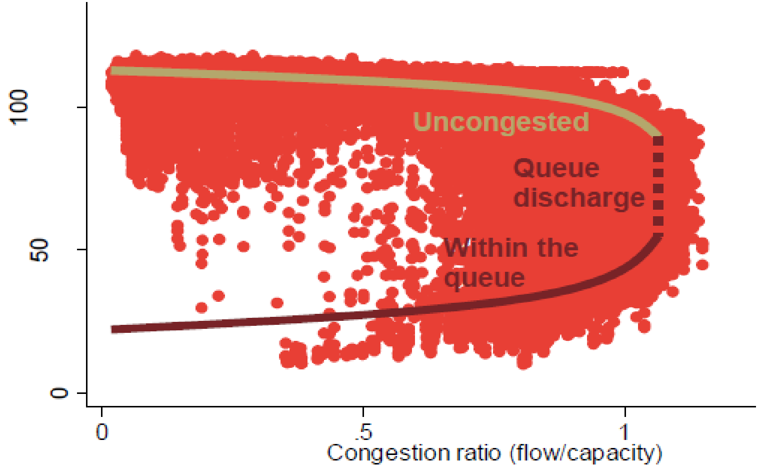
Computing

- Somebody else's problem to improve hardware or software, when you hit computational limits, either because of number of pieces of information or what you want to do with them.
- The end of Moore's Law, that computer power doubles every 18 months/two years?
- Computing needs depend on what you are trying to do, e.g. in finance applications
  - Description/diagnosis, identify patterns in financial data
  - Prediction, use the patterns and non-traditional data sources to forecast
  - Prescription, use the forecasts to develop a strategy
  - Automated Implementation of that strategy, algorithmic trading in real time
- Cannot do all the elements in real time.

Structured data

- Number of pieces of information, not a big problem if structured and you do not have to analyse in real time
- Roads project: data every 15 minutes, for 3 years (≈100k, for 2500 road sections with missing data (≈200m) for speed on the section and a set of 30 other variables (how many lanes working, whether there was an accident, weather)
- Just have to handle familiar problems on an industrial rather than hand-crafted scale:
- Example non-linearity,
  - When a road is uncongested, everything can speed along and it does not make a lot of difference adding a few more cars, then as it hits capacity it clogs up and speed drops sharply, going to zero when flow is also zero.
  - Graph shows theoretical relationship and data for a particular road section.

Figure 2. The relation between flow and speed (delays) is complex as show by the observed Flow-Speed curve on a road section between January 2007 and January 2009



Unstructured Data?

- What is data? Pictures, text, Need to put it in digital form, then in numerical form.
- Data is available in unstructured form, the first challenge is to structure it.
- Structuring means different things in different contexts depending on the purpose.
- Represent the data in numerical or graphical form, create a model, map to conclusions

Machine learning algorithms

- Supervised learning has an objective/loss function to evaluate the answer: does it predict? is it a cat?
- Unsupervised learning just looks for patterns (principal components/cluster analysis, latent Dirichlet allocation, latent topics) which may be given an interpretation
- Unsupervised learning from unstructured data remains a challenge.
- Cases where the supervised machine can evaluate its performance itself (e.g. winning at Chess, investing) versus cases where there are test data with the "right answer", problem of biased test data, e.g. all the faces in the test data are white men
- The data.may reflect prejudice already and the machine learns that
- Machine may not be able to explain why it made the decision
- In economics/finance supervised learning typically seeks functions that predict well out of sample.

Process

- Represent raw data d numerical form c
- map c to predictions $\widehat{v}$ of unknown outcomes v, (Supervised we observe v unsupervised we do not)
  - dictionary methods (list of policy related words) may be misleading out of context, finance related words different from general use
  - regression type $E(v \mid c)$
  - $p(c \mid v)$ (flu searches given flu)
- use $\widehat{v}$ for some purpose,
- Train the computer to read company reports very quickly as soon as they are released and trade on the basis of how numbers compare with expectations and linguistic features, how many times impairment is mentioned.

Number of explanatory variables >> Number of observations

- High dimensional environments but believe the models are sparse, large number of potential genes but believe only a few matter.
- More explanatory variables than observations: forms of penalised linear regression, e.g L2 costs ridge, L1 costs lasso, mixed L1 L2 elastic net etc. Minimise

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}\left|\beta_j\right|$$

- $\eta_i = \alpha + x_i'\beta$, $x_i = g(c_i)$, $E(v \mid x) = f(\eta_i)$ for some link function $f()$, e.g. logistic
- Support Vector machines and neural networks can tend to overfit and be difficult to tune. Deep neural nets with more layers and fewer nodes seem to work better?

Main challenge in economics and social sciences: Non stationary data?

- A cat stays a cat, but human behaviour changes, e.g. google flu search predictions.
- This is a big problem for cross validation methods, which are widely used in machine learning.
- Goodharts Law: any well established statistical relationship breaks down once it is used for policy.
- Lots of examples in finance and commerce: where profitable opportunities get competed away.
- Speed with which you can learn matters.
- Pandemic an extreme example of non-stationarity. FT Headline "Lowest level of output for 300 years." since 1709 Great Frost.

VARS and the Curse of dimensionality

- VAR(p)

$$\mathbf{y}_t = \mathbf{A}_0\mathbf{D}_t + \sum_{j=1}^{p} \mathbf{A}_j\mathbf{y}_{t-j} + \mathbf{u}_t$$

  $\mathbf{y}_t$ is $m \times 1$, $\mathbf{D}_t$ is a $k \times 1$ vector of deterministic elements, intercept, trend, seasonals.

- Each equation has $k + pm$ coefficients, gets big quickly. Quarterly seasonal data with trend $k = 5$, $p = 4$, $m = 6$, each equation has 31 parameters.
- May want m to be large, and often larger than T.
- Global VAR $T = 100$, $m \approx 150$ 5 variables for 30 countries.
- Road travel model, $\mathbf{y}_t$ travel time on 2500 sections of roads (e.g. between motorway junctions).

Ways of dealing with the curse of dimensionality

.

- Shrink the variable space, replace large number of variables by a few,
  - GVAR estimates national models with global averages as exogenous variables than can solve the whole system
  - factor models
- Shrink the parameter space:
  - Large Bayesian VARs, often use Minnesota priors that all the variables are random walks,
  - Lasso penalises non-zero parameters and so produces sparse models with lots of zeros.
- Covariance matrix is $m \times m$, $\mathbf{\Omega} = E(\mathbf{u}_t \mathbf{u}_t')$ but estimate of $\mathbf{\Omega}$, $\widehat{\mathbf{\Omega}}$, not positive definite and no inverse if $m > T$.
  - Shrink,to a diagonal matrix $\mathbf{\Lambda} = \mathbf{diag}\widehat{\mathbf{\Omega}}$ (which is always positive definite) $\widetilde{\mathbf{\Omega}} = (1-\lambda)\widehat{\mathbf{\Omega}} + \lambda\mathbf{\Lambda}$,where λ is large enough to make it positive definite.
  - Make it sparse by setting small elements of the correlation matrix $\mathbf{R}$ to zero.

Factor models 1

The term factor is used in lots of different contexts. Here it means that an observed series is driven by some unobserved components. Examples:

- Decomposition of time series into unobserved factors labelled trend, cycle, seasonals etc.as in "structural time series" models of Andrew Harvey.
- Identification of more theoretical underlying unobserved factors
  - developed most extensively in psychometrics, where the x_{it} are answers to a variety of questions by a sample of people. The underlying factors are aspects of personality, e.g. OCEAN: openness, conscientiousness, extroversion, agreeableness, neuroticism.
  - It has also been used in economics for unobserved variables like: development, natural rates, permanent components, core inflation, etc.

Factor models 2

- Used to measure the dimension of the independent variation in a set of data, e.g. how many factors are needed to account for most of the variation in x_{it}. If all the variables tend to move together a few factors will do.
- For I(1) series these dimensions may be the stochastic trends. and. replace the large number of possible x_{it} by a few f_{jt} which contain most of the information in the x_{it}. This may reduce omitted variable problems.
- May be relatively atheoretical.
- Choice depends on number of observed series, N; number of unobserved factors, r. and number of time periods, T.

Estimation Methods

- Univariate ($N = 1$) filters (e.g. the Hodrick-Prescott filter for trends).
- Multivariate ($N > 1$) but small, filters such as the Kalman filter used to estimate unobserved-component models.
- Multivariate ($N >> 1$) but large Principal component, PC,/Factor Analysis.
- 3 Pass Least Squares uses both cross-section and time series dimensions.
- Multivariate judgemental approaches, e.g. NBER cycle dating based on many series.
- Using a priori weighted averages of the variables.
- Deriving estimates from a model, e.g. Beveridge Nelson decompositions which treat the unobserved variable as the long-horizon forecast from a model.

Structural Time-Series Models

- These use the Kalman Filter to achieve this decomposition: Andrew Harvey *Forecasting, structural time series models and the Kalman filter*, CUP 1989 and program STAMP part of Oxmetrics suite.
- The unobserved components are the trend, τ_t, cycle c_t

$$\begin{aligned} y_t &= \tau_t + c_t + \omega_t \\ \tau_t &= \tau_{t-1} + \beta_t + \eta_t \\ \beta_t &= \beta_{t-1} + v_t \\ \phi(L)c_t &= \varepsilon_t \end{aligned}$$

- Shocks: $(\omega_t, \eta_t, v_t, \varepsilon_t)$ mean zero normally distributed, independent of each other and over time. Only small changes not big jumps or outliers.
- Deterministic trend: $\beta_t = \beta$, $E(\eta_t^2) = 0$. $\phi(L)$ polynomial in lag operator $\left(1 - \phi_1 L - \phi_2 L^2 ... - \phi_p L^p\right)$

State Space form

- Following Ghysels and Marcellino we can write it

$$\begin{aligned}
\alpha_t &= T\alpha_{t-1} + R\eta_t \\
y_t &= Z\alpha_t + S\xi_t \\
\begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix} &\sim \underset{iid}{\sim} N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} Q & 0 \\ 0 & H \end{bmatrix} \right),
\end{aligned}$$

- α_t is a vector containing the unobservable state variables, and the first equation describes the evolution of them, y_t are the observed data and the second, transition or measurement, equation models them as function of the state variables.
- It is often very useful to cast models into this form because they can then be estimated using the Kalman filter.

State Space form uses

- Kalman Filter is a recursive method that gives linear forecasts for α_t and y_t and the model likelihood (given normality). Can be used for prediction, given information up to t to forecast $t+1$, or for smoothing, which uses all the information in the sample.
- Applications
  - Unobserved components and with cycle and seaonality
  - ARIMA models can be written in this way.
  - Time varying parameter models can be written in this way.
  - Stochastic volatility models.
  - Models with missing data.
  - dynamic factor models for a small number of variables G&M give an example with 4 growth rates.
  - DSGE models for estimation and solution.

Issues

- Unobserved component models for small N factors tend to put more parametric structure on the factors, which PCs do not. Principal Components, PC, can be appropriate for large N small r.
- Long history of PCs. In the early days it was not clear whether the errors in variables model (observed data generated by unobserved factors) or the errors in equation model was appropriate and both models were used, for instance
- Stone (1947) used PCs to show that most of the variation in a large number of national accounts series could be accounted for by three factors, which could be interpreted as trend, cycle and rate of change of the cycle.
- Given a $T \times N$ data matrix, X with typical element x_{it}, (i usually variable here)

Factor models

- Factor model takes the form

$$x_{it} = \lambda_i' f_t + e_{it} \tag{1}$$

  - f_t are a $r \times 1$ vector of common factors, with $r < N$; variable reduction
  - λ_i is a $r \times 1$ vector of factor loadings and
  - e_{it} is an idiosyncratic component.
- Strictly Factor models are statistical models, where one makes some assumption about the distribution of e_{it} and estimates by maximum likelihood but in practice one uses a mathematical procedure called Principal Components.
- The PCs are $T \times 1$ vectors linear combinations of X that have maximal variance and are orthogonal. $F = XA$
- The X matrix is often first standardised, $X'X$ is then the correlation matrix. The PCs are not invariant to scaling.

Principal Components: estimation

- To obtain the first PC we construct a $T \times 1$ vector $f_1 = Xa_1$, such that $f_1' f_1 = a_1' X' X a_1$ is maximised, subject to $a_1' a_1 = 1$.
- The problem is then a Langrangian

$$L = a_1' X' X a_1 - 2\phi_1 (a_1' a_1 - 1)$$

$$\frac{\partial L}{\partial a_1} = 2X' X a_1 - 2\phi_1 a_1 = 0$$

$$X' X a_1 = \phi_1 a_1$$

- So a_1 is the first eigenvector of $X'X$, (the one corresponding to the largest eigenvalue, ϕ_1); .
- The second PC $f_2 = Xa_2$ is the linear combination which has the second largest variance, subject to being uncorrelated with a_1 i.e. $a_2' a_1 = 0$, so a_2 is the second eigenvector.

Properties

- If X is of full rank, there are N PCs.
- Note $AA' = I_N$ so $A' = A^{-1}$. Stacking

$$X'XA = A\Phi.$$

and

$$\begin{aligned} A'X'XA &= \Phi \\ F'F &= \Phi \end{aligned}$$

- The eigenvalues measure the proportion of the variation in X that each principal component explains: $\phi_i / \sum \phi_i$. If the data are standardised $\sum \phi_i = N$, the total variance.
- Forming the PCs is a purely mathematical operation replacing the $T \times N$ matrix X, by the $T \times N$ matrix F.

Factor Models

- We define the PCs as $F = XA$, but we can also write $X = FA'$ defining X in terms of the PCs. Usually we want to reduce the number of PCs that we use, to reduce the dimensionality of the problem so we can write it

$$X = F_1 A_1' + F_2 A_2'$$

  where the $T \times r$ matrix F_1 contains the $r < N$ largest PCs, the $r \times N$ matrix A_1' contains the first r eigenvectors corresponding to the largest eigenvalues. We treat F_1 as the common factors corresponding to the f_t, and $F_2 A_2$ as the idiosyncratic factors.
- We will use F for both all the N PCs and for relevant subset of r PCs.
- We will focus on static factor models, which extract the PCs of the covariance matrix. There are also dynamic factor models which extract the PCs of the long-run covariance matrix or spectral density matrix.

Example

- Cochrane gives the eigenvalue decomposition of the covariance matrix of zero coupon bond yields of different maturities.

TABLE A6

BOND YIELD FACTOR ANALYSIS

σ	1	2	3	4	5
6.36	0.45	0.45	0.45	0.44	0.44
0.61	-0.75	-0.21	0.12	0.36	0.50
0.10	0.47	-0.62	-0.41	0.11	0.46
0.08	0.10	-0.49	0.39	0.55	-0.55
0.07	0.07	-0.36	0.68	-0.60	0.21

σ, square root of the eigenvalues. rows the eigenvectors corresponding to 1-5 year yields.

- The first eigenvector, accounts for the bulk of the variation, has roughly equal weights and gives level, 2nd gives slope, 3rd curvature.

Interpretation

- This is a rare case where there is an easy interpretation, but first PC is often close to the mean of the data.
- Identification? For any non-singular matrix P then $(FP^{-1})(P\Lambda) = \widetilde{F}\widetilde{\Lambda}$ is observationally equivalent to $F\Lambda$. Interpretation of PCs can be a major problem if you do not have a way to identify them. Eigenvector restrictions, unit length and orthogonal, may not reflect the economics.
- For forecasting may not need to identify factors.
- Choice of N? How many variables to include in X raises issues often r rises with N, since additional variables contain different factors.
- Estimated or imposed weights? Imposed weights biased but no variance, estimated weights may be unbiased but have large variance. So imposed weights may have smaller MSE, be easier to interpret and may reflect purpose better.

Choice of r, the number of factors

- The choice of the number of factors is a major issue.
- Ad hoc rules like choose PCs corresponding to eigenvalues greater than one on standardised data.
- By construction factors minimise the unexplained variance

$$V(r) = (NT)^{-1} \sum_i \sum_t (X_{it} - \lambda_i' F_t)^2.$$

- Bai and Ng (2002) show that the number of factors r can be estimated consistently as $\min(N, T) \to \infty$ e.g. by minimising criteria like

$$\log(V(r)) + r \left(\frac{N+T}{NT} \right) \log \left(\frac{NT}{N+T} \right)$$

- This may not work well when N or T are small, tending to choose the max r allowed.

Restrictions on regression

- Using PCs can Reparameterise

$$y = X\beta + u$$

$$\begin{aligned} y &= XAA'\beta + u \\ y &= F\delta + u \end{aligned}$$

- Then reduce the number of PCs by writing it

$$\begin{aligned} y &= F_1\delta_1 + F_2\delta_2 + u \\ y &= XA_1A_1'\beta + XA_2A_2'\beta + u \end{aligned}$$

- Test $\delta_2 = 0$ if $N < T$ and if accepted use $y = F_1\delta_1 + e$.
- Note **F** chosen to explain X not y. Interpretation of restriction $A_2'\beta = 0$?

Explanation with dynamic factor models

- Usual procedure is to difference the data to make it stationary then extract factors and use them as extra variables in a VAR: factor augmented VAR, FAVAR This loses all the levels information.
- If one extracts I(1) factors, and estimates factor augmented ECMs, FECMs, one must ensure that dependent variables and factors cointegrate.
- Dynamic factor models use PCs of long-run covariance matrix. They are two sided filters but may be approximated by lags of static factors.

FAVAR 1

- Consider a $M \times 1$ vector of observed variables $\mathbf{Y}_t$, a $K \times 1$ vector of unobserved factors $\mathbf{F}_t$ with a VAR structure

$$\begin{pmatrix} \mathbf{F}_t \\ \mathbf{Y}_t \end{pmatrix} = \mathbf{A}(L) \begin{pmatrix} \mathbf{F}_{t-1} \\ \mathbf{Y}_{t-1} \end{pmatrix} + \mathbf{v}_t.$$

  where $\mathbf{A}(\mathbf{L})$ is a polynomial in the lag operator.
- The unobserved factors $\mathbf{F}_t$ are related to a $N \times 1$ vector $\mathbf{X}_t$, which contains a large number (BBE use $N = 120$, LM $N = 105$) of potentially relevant observed variables by

$$\mathbf{X}_t = \mathbf{\Lambda} \mathbf{F}_t + \mathbf{e}_t,$$

- The $\mathbf{F}_t$ are estimated as the principal components of the $\mathbf{X}_t$, which may include the $\mathbf{Y}_t$.

FAVAR 2

The advantage of FAVARs is that

- a small number of factors can account for a large proportion of the variance of the $\mathbf{X}_t$ and thus parsimoniously reduce omitted variable bias in the VAR ;
- the factor structure for $\mathbf{X}_t$ allows one to calculate impulse response functions for all the elements of $\mathbf{X}_t$ in response to a (structural) shock in $\mathbf{Y}_t$ transmitted through $\mathbf{F}_t$;
- the factors may be better measures of underlying theoretical variables such as economic activity than the observed proxies such as GDP or industrial production; but PCs can be difficult to interpret
- FAVARs may forecast better than standard VARs
- factor models can approximate infinite dimensional VARs,

Forecasting with FAVAR

.

- This is exactly the same as a standard VAR:

$$\begin{aligned}\mathbf{F}^f_{T+1,T} &= \mathbf{A}_{11}\mathbf{F}_T + \mathbf{A}_{12}\mathbf{Y}_T \\ \mathbf{Y}^f_{T+1,T} &= \mathbf{A}_{21}\mathbf{F}_T + \mathbf{A}_{22}\mathbf{Y}_T\end{aligned}$$

- with the advantage that you can also forecast the $\mathbf{X}_t$

$$\mathbf{X}^f_{T+1,T} = \mathbf{\Lambda}\mathbf{F}^f_{T+1,T}$$

- Run these forward in the usual way.

3 Pass Regression Filter 1

- Developed by Kelly & Pruit, See Ghysels and Marcellino.
- Works for nowcasting and forecasting (related to Partial Least Squares and Fama Macbeth). Model

$$\begin{aligned} y_{t+1} &= \beta_0 + \beta' F_t + \eta_{t+1} \\ z_t &= \lambda_0 + \Lambda F_t + \omega_t, \\ x_t &= \Phi_0 + \Phi F_t + \varepsilon_t. \end{aligned}$$

- y_t $t = 1, 2, ..., T$ is the target variable of interest $F_t = (f_t', g_t')$, are the $K = K_f + K_g$ factors driving all the variables, $\beta' = (\beta_f', 0)$ so y_t only depends on f_t; z_t is a small set of L proxies driven by the same factors as y_t, so $\Lambda = (\Lambda_f, 0)$ where Λ_f is non-singular, x_t is a large set of N variables driven by both (f_t', g_t')

3 Pass Regression Filter 2

- Pass 1: run a set of N time-series regressions of each element of x on z_t
$$x_{it} = \phi_{0i} + z_t^{'}\phi_i + \varepsilon_{it},$$
keep the L OLS estimates of the ϕ_i.
- Pass 2: run a set of T cross section regressions of x_{it} on the L ϕ_i for each t
$$x_{it} = \phi_{0t} + \widehat{\phi}_i F_t + \varepsilon_{it}$$
keep the OLS estimates $\widehat{F}_t$
- Pass 3: run a time series regression
$$y_{t+1} = \beta_0 + \beta' \widehat{F}_t + \eta_{t+1}.$$
Then use this to forecast.

3 Pass Regression Filter 3

- For one factor use the target variable y_t as the proxy z_t. target-proxy 3PRF
- For more than one factor they suggest either using theory based proxies or a procedure that can be automated described in Ghysels and Marcellino

GVAR

- GVAR is a very widely used structure,
  - Can estimate model unit by unit
  - but put all the unit specific models together using a weighting matrix like the spatial weighting matrix
  - to create an interdependent system which can analyse interactions between all the units.
- GVAR toolbox has data and matlab code: Vanessa Smith.
  - https://sites.google.com/site/gvarmodelling/gvar-toolbox
- GVAR Handbook and Pesaran's web pages have lots of other examples.

Example

- use as an example Dees et al. (2007) DdPS. They have for $N = 26$ countries, $m = 6$ variables, , so $mN = 156$, Given quarterly data $T = 100$, 1979Q4-2003Q4, unrestricted estimation clearly not possible.
- The long-run matrix $\boldsymbol{\Pi} = \boldsymbol{\alpha\beta}$ has 63 cointegrating vectors, both within and between countries.
- Specific example of a general problem, where we want to estimate an infinite order VAR and ask: What restrictions might let us do this?.

Use trade weights

- The GVAR approach, allows us to shrink the parameter space using a known weighting matrix,
- In the case of the GVAR the trade weights between country i and j.
- But it is a very general framework for constructing a large VAR, where direct estimation is impossible.
- Other weighting matrices that have been used include
  - input-output matrices for inter-industry linkages;
  - financial flow matrices;
  - networks of contacts for interaction effects;
  - neighborhood matrices for spatial models and spillover effects;
  - equal weights as in CCE estimator; etc.

VARX*

- Countries $i = 0, 1, 2, ..., N$, with country 0, say the US, as numeraire. For country i. VARX*(2,2) as

$$\begin{aligned} \mathbf{x}_{it} &= \mathbf{B}_{id}\mathbf{d}_t + \mathbf{B}_{i1}\mathbf{x}_{i,t-1} + \mathbf{B}_{i2}\mathbf{x}_{i,t-2} + \\ &\quad \mathbf{B}^*_{i0}\mathbf{x}^*_{it} + \mathbf{B}^*_{i1}\mathbf{x}^*_{i,t-1} + \mathbf{B}^*_{i2}\mathbf{x}^*_{i,t-2} + \mathbf{u}_{it}, \end{aligned}$$

- $\mathbf{x}_{it}$ is a $k_i \times 1$ vector of domestic variables,
- $\mathbf{x}^*_{it}$, a $k^*_i \times 1$ vector of foreign variables specific to country i, and $\mathbf{d}_t$ a $s \times 1$ vector say $(1, t, p^o_t)'$.

$$\mathbf{x}^*_{it} = \sum_{j=0}^{N} w_{ij}\mathbf{x}_{jt}, \text{ with } w_{ii} = 0,$$

w_{ij} share of country j in the trade i.

VECM

- VARX* can be estimated separately for each country, taking into account the possibility of cointegration between $\mathbf{x}_{it}$ and $\mathbf{x}^*_{it}$ and within $\mathbf{x}_{it}$.
- Write as VECM

$$\begin{aligned}\Delta\mathbf{x}_{it} &= \mathbf{B}_{id}\mathbf{d}_t - \mathbf{\Pi}_i\mathbf{z}_{i,t-1} + \\ &\quad \mathbf{B}^*_{i0}\Delta\mathbf{x}^*_{it} + \mathbf{\Gamma}_i\Delta\mathbf{z}_{i,t-1} + \mathbf{u}_{it},\end{aligned}$$

- Where $\mathbf{z}_{it} = (\mathbf{x}'_{it}, \mathbf{x}^{*\prime}_{it})'$ and $\mathbf{\Pi}_i = \boldsymbol{\alpha}_i\boldsymbol{\beta}'_i$,
- Can write it

$$\begin{aligned}\Delta\mathbf{x}_{it} &= -\boldsymbol{\alpha}_i\boldsymbol{\beta}'_i\left(\mathbf{z}_{i,t-1} - \mathbf{\Upsilon}_i\mathbf{d}_{t-1}\right) + \\ &\quad \mathbf{B}^*_{i0}\Delta\mathbf{x}^*_{it} + \mathbf{\Gamma}_i\Delta\mathbf{z}_{i,t-1} + \mathbf{\Pi}_i\mathbf{\Upsilon}_i\Delta\mathbf{d}_t + \mathbf{u}_{it},\end{aligned}$$

Feedbacks and weak exogeneity

- The r_i error correction or disequilibrium terms are

$$\boldsymbol{\xi}_{it} = \boldsymbol{\beta}_i' \mathbf{z}_{it} - \boldsymbol{\beta}_i' \boldsymbol{\Upsilon}_i \mathbf{d}_t = \boldsymbol{\beta}_{ix}' \mathbf{x}_{it} + \boldsymbol{\beta}_{ix*}' \mathbf{x}_{it}^* + \boldsymbol{\gamma}_i' \mathbf{d}_t,$$

- Weak exogeneity implies $\boldsymbol{\alpha}_i^* = 0$ in

$$\Delta \mathbf{x}_{it}^* = \mathbf{c}_{i0}^* + \boldsymbol{\alpha}_i^* \boldsymbol{\xi}_{i,t-1} + \boldsymbol{\Gamma}_i^* \Delta \mathbf{z}_{i,t-1} + \mathbf{u}_{it}^*.$$

- This can be tested.

Variables

- logarithm of real output, y_{it};
- inflation, $\pi_{it} = p_{it} - p_{it-1}$,
- exchange rate variable, which is defined as $e_{it} - p_{it}$,
- a short interest rate, r_{it}^S
- a long interest rate, r_{it}^L
- logarithm of real equity prices, q_{it}.

$\mathbf{x}_{it}$ is the vector of these variables for country i.
$\mathbf{x}_t = (\mathbf{x}'_{0t}, \mathbf{x}'_{1t}, ..., \mathbf{x}'_{Nt})'$ is the vector of all the endogenous variables, for all countries.

Stacking the system

Write VARX*

$$\begin{aligned} \mathbf{x}_{it} - \mathbf{B}^*_{i0}\mathbf{x}^*_{it} &= \mathbf{B}_{id}\mathbf{d}_t + \left(\mathbf{B}_{i1}\mathbf{x}_{i,t-1} + \mathbf{B}^*_{i1}\mathbf{x}^*_{i,t-1}\right) + \\ &\quad \left(\mathbf{B}_{i2}\mathbf{x}_{i,t-2} + \mathbf{B}^*_{i2}\mathbf{x}^*_{i,t-2}\right) + \mathbf{u}_{it}, \end{aligned}$$

$$\mathcal{A}_{i0}\mathbf{z}_{it} = \mathbf{h}_{i0} + \mathbf{h}_{i1}t + \mathcal{A}_{i1}\mathbf{z}_{it-1} + \mathcal{A}_{i2}\mathbf{z}_{it-2} + \mathbf{u}_{it},$$

for $i = 0, 1, 2, ..., N$ using $\mathbf{z}_{it} = (\mathbf{x}'_{it}, \mathbf{x}^{*\prime}_{it})'$ and using $\mathbf{z}_{it} = \mathbf{W}_i\mathbf{x}_t$,

$$\mathcal{A}_{i0}\mathbf{W}_i\mathbf{x}_t = \mathbf{h}_{i0} + \mathbf{h}_{i1}t + \mathcal{A}_{i1}\mathbf{W}_i\mathbf{x}_{t-1} + \mathcal{A}_{i2}\mathbf{W}_i\mathbf{x}_{t-2} + \mathbf{u}_{it},$$

Solving the system

- VARX*

$$\mathcal{A}_{i0}\mathbf{W}_i\mathbf{x}_t = \mathbf{h}_{i0} + \mathbf{h}_{i1}t + \mathcal{A}_{i1}\mathbf{W}_i\mathbf{x}_{t-1} + \mathcal{A}_{i2}\mathbf{W}_i\mathbf{x}_{t-2} + \mathbf{u}_{it},$$

  for each country $i = 0, 1, 2, ..., N$.
- Can be stacked to yield model for $\mathbf{x_t}$

$$\mathbf{H}_0\mathbf{x}_t = \mathbf{h}_0 + \mathbf{h}_1\mathbf{t} + \mathbf{H}_1\mathbf{x}_{t-1} + \mathbf{H}_2\mathbf{x}_{t-2} + \mathbf{u}_t,$$

  $\mathbf{H}_0$ is a known non-singular matrix that depends on the trade weights and parameter estimates..
- Multiply by $\mathbf{H}_0^{-1}$ to get the GVAR

$$\mathbf{x}_t = \mathbf{a}_0 + \mathbf{a}_1\mathbf{t} + \mathbf{G}_1\mathbf{x}_{t-1} + \mathbf{G}_2\mathbf{x}_{t-2} + \mathbf{v}_t,$$

  $\mathbf{G}_1 = \mathbf{H}_0^{-1}\mathbf{H}_1$, etc

Dimensions

- 134 endogenous variables
- 71 stochastic trends
- 63 cointegrating relations
- long run forcing assumption rejected in 5 out of 153 cases
- structural instability is found primarily in the error variances (47% of the equations - clustered in the period 1985-1992).
- between country interdependence: through $\mathbf{x}_{it}^*$ variables (large); oil prices, and
- error covariances, $\boldsymbol{\Sigma} = \mathbf{E}(\mathbf{v}_t \mathbf{v}_t')$ $\widehat{\boldsymbol{\Sigma}}_{ij} = \sum_t \widehat{\mathbf{v}}_{it} \widehat{\mathbf{v}}_{jt}' / T$. Correlations small except for real exchange rate equations.

Steady states

- Steady states are long-horizon forecasts, can be obtained from GVAR, provide estimates of multivariate Beveridge-Nelson trends, which incorporate stochastic trends and cointegrating vectors.
- Have advantages over purely statistical measures like the HP filter.
- Univariate BN trends are not smooth enough, not the case with multivariate BN trends
- Then can calculate deviations from steady states (e.g. output gaps), which will be $I(0)$ by construction and use them in DSGE type structural modelling.

Nowcasting

- Your GDP forecast will start out from this quarters GDP, but you dont know this because the quarterly figure is published with a lag.
- However, you have higher frequency data on unemployment and employment monthly, retail sales, survey measures like PMI, purchasing managers index, and maybe higher frequency indicators like electricity consumption, electronic bank payments data..
- Nowcasting models are
  - mixed frequency models which forecast quarterly GDP from higher frequency series
  - identify factors from big data sets
  - with a midel selection procedure to choose which of the large number of potential factors are useful
  - procedurs to switch between models depending on other indicators.

Bridge Equations

- These link (bridge) high frequency series like monthly employment and industrial production, $x^H_{t,m}$ to lower frequency series like quarterly GDP y^L_t.
- We can aggregate the higher frequency series, say monthly, to quarterly $x^L_t = \sum_{m=1}^3 x_{t,m}$, other ways of aggregating may be used.
- We can the estimate

$$y^L_t = a + bx^L_t + u^L_t.$$

  though usually a dynamic model would be used.
- We now want to predict the first out of sample observation

$$y^L_{T+1} = \widehat{a} + \widehat{b}x^L_{T+1}$$

Bridge Equation forecasts

- Suppose we have one months real data $x^H_{T+1,1}$, we then supplement it with forecasts of months 2, $\widehat{x}^H_{T+1,2}$ and 3 $\widehat{x}^H_{T+1,3}$ to get

$$\widehat{x}^L_{T+1,1} = x^H_{T+1,1} + \widehat{x}^H_{T+1,2} + \widehat{x}^H_{T+1,3}$$

  with forecast

$$\widehat{y}^L_{T+1,1} = \widehat{a} + \widehat{b}\widehat{x}^L_{T+1,1}.$$

- When month 2's data becomes available we can replace the forecast for month 2 with the actual and update our forecast of month 3

$$\widehat{x}^L_{T+1,2} = x^H_{T+1,1} + x^H_{T+1,2} + \widehat{x}^H_{T+1,3}$$

  and get a new estimate of GDP

$$\widehat{y}^L_{T+1,2} = \widehat{a} + \widehat{b}\widehat{x}^L_{T+1,2}.$$

Questions to Discuss

- Weather Forecasting has improved massively in the last 60 years economic forecasting has not improved at all. What's the difference?
- David Hendry noted that when weather forecasts fail the government buys the Met Office a new £1.2bn computer; when economic forecasts fail the government cuts economics research funding. Would a .£1.2bn computer improve economic forecasting?
- Tim Harford in an article on Moodle argues that Economists can learn a lot from meteorologists. Do you agree?
- Suppose.that you have T observtions on m interconnected variables that you want to predict and $m >> T$. Discuss the advantages and disadvantages of a number of alternatives procedures that you might adopt.