Overview of Recent Advances in Macroeconomic Forecasting

Background to this Lecture

- I was involved with a team of Bayesians led by Gael Martin to write a survey paper: "Bayesian Forecasting in Economics and Finance: A Modern Review" for the International Journal of Forecasting
- I was involved with writing the section on macroeconomic forecasting
- Preliminary draft of our section available on course website and this lecture based on it
- The aim of this lecture is to offer an overview of which models are commonly used by Bayesians for macro forecasting
- ... and why
- ... and what the key issues are
- In later lectures we will get into details of most of these models

Macroeconomists Have Lots of Data

- For many countries can easily get data for over 100 variables
- With globalization may want to work with several countries (so even more variables)
- US: FRED-MD: A Monthly Database for Macroeconomic Research (Federal Reserve Bank of St. Louis)
- 134 variables (output, prices, consumption, interest rates, stock prices, money, housing, unemployment, wages, etc. etc. etc.)
- Why work with all of them?
- When forecasting (e.g. inflation, GDP growth, unemployment) the more information the better
- When estimating a model want to avoid omitted variables bias
- If other variables have important explanatory power and you omit them, model is mis-specified
- We want large models which jointly model many variables

Why Bayesian?

- Large-scale models are prone to overfitting (many parameters to estimate with relatively few observations) which leads to poor forecasts
- But Bayesian methods can handle many parameters since:
- Prior shrinkage helps with over-fitting
- Modern Bayesian methods can deal with issues of specification uncertainty that arise in large models

Why Bayesian?

- Why not do hypothesis testing to reduce K?
- Pre-test problem (also called multiple testing or multiple comparisons problem)
- Last lecture had cross-country growth regression example with $K = 41$
- There are 41 different restricted regressions which drop one of the explanatory variables
- There are $\frac{K(K-1)}{2}$ restricted regressions with drop two of the explanatory variables
- etc. etc. etc.
- In total there are $2^K = 2,199,023,255,552$ possible regression models involving some combination of the explanatory variables
- Jargon: this is the model space
- Which one to choose?

Why Bayesian?

- Sequential hypothesis testing methods often used in smaller problems
- Let us suppose you can come up with a sequence of hypothesis tests to navigate through your huge model space
- E.g. do a hypothesis test to decide whether to drop a variable, then do a second hypothesis test using the restricted regression
- But significance levels no longer valid (or must be adjusted) when more than one test is done
- E.g. one t-test using standard critical value has 5% level of significance. But if you do two t-tests sequentially second one no longer has 5% level of significance
- Maybe minor issue in small data problems, but with Fat Data problems number of sequential hypothesis tests may be HUGE, true level of significance vastly different from nominal one (or necessary adjustments become huge)
- Bottom line: not easy to do hypothesis testing to select a more parsimonious model

A Digression on DSGE Models

- Dynamic stochastic general equilibrium (DSGE) models popular with macroeconomists
- They use macroeconomic theory to define a likelihood function
- They are state space models (cover later in the course)
- Commonly estimated using Bayesian methods (Metropolis-Hastings)
- Mostly used for structural economic analysis, occasionally used for forecasting
- This course will not say much about DSGE models
- We will focus on reduced form multivariate time series models which are more commonly used for forecasting

Macroeconomic Forecasting Models

- General class of models we work with:
- $\boldsymbol{y}_t$ is M vector of variables to be forecast, depends on K-dimensional vector $\boldsymbol{x}_t$ (K is often large)
- $$\boldsymbol{y}_t = g(\boldsymbol{x}_t) + \varepsilon_t$$
- $g : \mathbb{R}^K \rightarrow \mathbb{R}^M$ is some function
- ε_t is $\mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t)$
- Note: Normal error assumption not restrictive since can use hierarchical priors $\boldsymbol{\Sigma}_t$ to get more flexible error distributions
- E.g. one kind of mixture of Normals leads to Student-t errors, other mixtures lead to Bayesian nonparametric treatments, etc.
- Almost all Bayesian macroeconomic forecasting models fall in this general class but involve different choices for $\boldsymbol{y}_t$, $\boldsymbol{x}_t$, g, etc.

The Vector Autoregressive Model (VAR)

- $\boldsymbol{y}_t$ is vector of dependent variables
- $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})'$ contains p lags of $\boldsymbol{y}_t$
- $g(\boldsymbol{x}_t) = \boldsymbol{A}\boldsymbol{x}_t$ is a linear function with $M \times K (= Mp)$ coefficient matrix $\boldsymbol{A}$
- $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$ is constant over time
- This is standard homoskedastic VAR commonly used for forecasting for many years

Factor Models

- $\boldsymbol{y}_t$ is vector of dependent variables
- $\boldsymbol{x}_t = \boldsymbol{f}_t$ with $\boldsymbol{f}_t$ denoting a set of $Q \ll M$ latent factors
- $g(\boldsymbol{f}_t) = \boldsymbol{\Lambda} \boldsymbol{f}_t$ is linear
- $\boldsymbol{\Lambda}$ being an $M \times Q$ matrix of factor loadings
- $\boldsymbol{f}_t$ evolves according to a VAR
- This is the dynamic factor model (DFM)
- Combining DFM with VAR leads to Factor Augrmented VAR (FAVAR)

Single Equation Models

- Sometimes macroeconomic forecasters use single equation methods:
- $\boldsymbol{y}_t$ is a single variable (e.g. inflation), g is linear and $\boldsymbol{x}_t$ are predictors (e.g. based on the Phillips curve)
- This leads simply to regression methods
- For inflation forecasting another model is also popular: Unobserved components stochastic volatility (UCSV) model
- $g(\boldsymbol{x}_t) = \alpha_t$ where α_t is a latent variable which follows a random walk (UC)
- $\boldsymbol{\Sigma}_t$ follows a stochastic volatility process (SV)

Parameter Variation and Nonlinearity

- Traditionally, VARs and DFMs have been linear and homoskedastic.
- But empirical evidence in most macroeconomic data sets of parameter change, both in the conditional mean and the conditional variance
- Accomodating this leads to various forms for g and $\boldsymbol{\Sigma}_t$.
- For $\boldsymbol{\Sigma}_t$ SV is very popular.
- For conditional mean, one parametric form for g leads to time-varying parameter VARs (TVP-VARs)
- Another to Markov switching VARs, etc.
- Alternatively, assume g is unknown and use Bayesian nonparametric methods to uncover its form

Bayesian Forecasting Using These Models

- This general framework defines a class of likelihood functions.
- Bayesians multiply likelihood function by an appropriate prior to produce a posterior which can be used to produce the predictive density
- The choice of prior and computational method used for posterior and predictive inference will be case specific, but a few general issues worth stressing
- Priors:
- Choice of prior matters much more in models which have large number of parameters relative to the number observations than in models with fewer parameters
- Hence, lots of recent research on priors for VARs, less for models such as the UCSV or the DFM

Bayesian Forecasting Using These Models

- Computation:
- Linear homoskedastic models with conjugate priors lead to analytical formulae for the posterior and the 1-step ahead predictive density
- For all other cases, MCMC methods are available
- If g is linear, these are standard and familiar involving Bayesian algorithms for regression and/or state space models.
- But MCMC is slow in models involving large numbers of parameters (such as VARs) or large numbers of latent states (such as TVP-VARs)
- Thus, much recent research on developing improved MCMC algorithms or approximate methods for speeding up computation.

Bayesian Forecasting Using These Models

- Forecast combination:
- But forecast performance often improved by combining forecasts from multiple models
- There are many different Bayesian methods for forecast combination (e.g. BMA)
- Moving Beyond Point Forecasts:
- Traditionally, point forecasts reported
- Increasing interest in forecast uncertainty
- Also interest in tail risk
- Bayesian methods produce predictive density which can be used for either

Recent Developments: Large VARs

- Large VARs beat DFMs in many forecasting horse races
- Lots of parameters: each equation has $K = Mp$ where p is lag length
- In total at least pM^2 VAR coefficients
- Influential early application had $M = 100$, $p = 12$
- Computation can be time consuming
- Technical note: key computational bottleneck involves the posterior covariance matrix of the VAR coefficients
- Matrix manipulations involving high dimensional matrices slow and sometimes unstable
- One trick which helps: transform reduced form VAR into structural VAR with diagonal error covariance and then do estimation one equation at time (each with much smaller covariance matrix)
- VAR is really just a series of regressions so if you know Bayesian methods for regressions, Bayesian methods for VARs are easy

Recent Developments: New Priors for VARs

- Need prior shrinkage for large VARs
- Several standard priors (conjugate prior, Minnesota prior, etc.) have been used for decades
- Require subjective choice of prior hyperparameters (e.g. degree of shrinkage imposed)
- Recently much interest in more objective methods (e.g. estimating appropriate degree of shrinkage)
- One popular class (associated with machine learning methods): global-local shrinkage priors

Recent Developments: Global-local Shrinkage Priors

- Let a_j be the j^{th} VAR coefficient
- $$a_j \sim \mathcal{N}(0, \psi_j \lambda), \quad \psi_j \sim f_1, \quad \lambda \sim f_2,$$
- λ controls global shrinkage (common to all coefficients)
- ψ_j controls local shrinkage (specific to the j^{th} coefficient)
- λ and ψ_j are parameters – optimal degree of shrinkage for each coeff. estimated from data
- f_1 and f_2 are mixing densities and different choices lead to different priors with different properties
- Lasso, Horseshoe, Stochastic search variable selection (SSVS), Dirichlet-LaPlace, etc. etc.
- Setting up MCMC algorithms is easy: conditional on λ and ψ_j just a Normal prior (just need new Gibbs steps for drawing from them)

Recent Developments: Parameter Change and Nonlinearity

- Adding SV typically improves forecast performance, particularly when interest is in density forecasts
- VAR-SVs are the workhorse model for many Bayesian macroeconomic forecasters
- I have found allowing for changes in the VAR coefficients to be less important, but it is occasionally useful
- TVP-VAR-SV allows for random walk (or AR) change in VAR coeffs.
- This is a parametric approach
- Recently, interest in nonparametric approaches: regression trees, Gaussian processes, infinite mixtures, etc.
- Nonparametric techniques work well in unstable times (e.g. covid-19 outliers)

Summary

- This lecture describes class of models used by Bayesians for macroeconomic forecasting
- And relevant issues (prior choice and computational challenges in large models)
- Many more specialised cases within this general framework
- Restricted versions of VARs: Vector Error Correction Models (VECMs) which impose cointegrating restrications, global VARs, spatial VARs
- This lecture has focussed on forecasting, for nowcasting can use mixed frequency VARs
- In following lectures, will provide Bayesian details for many of the models discussed in this lecture