Nowcasting by the BSTS-U-MIDAS Model

Jun Duan

Supervised by: David Giles

September 14, 2015

University of Victoria
Economics

Outline

BSTS-U-MIDAS Model

Proposition: a BSTS-U-MIDAS model (Bayesian Structural Time Series-Unrestricted-Mixed-Data Sampling model)

- structural time series model (STM).
- MIDAS model.
- spike-and-slab regression.
- Bayesian model averaging (BMA).

Application: forecast quarterly GDP for Canada. More accurate than ARIMA and Boosting models. Capture the structural break of the 2008-2009 crisis.

Forecasting with Mixed Frequency Data

Problems: Using high frequency data for forecasting or nowcasting low frequency data.

- the mixed frequency problem.
- the unbalanced data problem (missing observations, ragged edge data).
- the high dimensionality (fat regression, parameter proliferation) problem.

Literature Review

Research Question

Utilize the high-frequency data to improve forecasts of low-frequency macroeconomics variables.

1. Aggregation and interpolation.
2. MIDAS models.
3. Factor models.
4. Statistical learning Approach.

MIDAS Model

Directly introduces high frequency data into the equation (Ghysels, 2004).

$$\text{Time: } \begin{bmatrix} y_{2nd\ quarter} & & ; & y_{1st\ quarter} & & ; \dots \\ x_{June} & x_{May} & x_{Apr}; & x_{Mar} & x_{Feb} & x_{Jan}; \dots \end{bmatrix}$$

$$\text{Alignment: } \begin{bmatrix} y_{2nd\ quarter} & = & \beta_1 x_{June} & +\beta_2 x_{May} & +\beta_3 x_{Apr} \\ y_{1st\ quarter} & = & \beta_1 x_{Mar} & +\beta_2 x_{Feb} & +\beta_3 x_{Jan} \\ \dots & = & \dots & \dots & \dots \end{bmatrix}$$

- Advantage is easy to incorporate leading variables.
- The missing values are filled with NA or forecasts of ARIMA process.
- x_i is quarterly skip-sampled.

Factor Models

A few factors summarize many variables. Principal components regression; dynamic factor model. (Banbura, 2013).

$$X_t = \lambda(L) f_t + e_t$$

$$f_t = \Psi(L) f_{t-1} + \eta_t$$

Solved in a state-space form.
High frequency f_t is aggregated to low frequency $\mathbf{f_t}$.

$$y_t = \beta_0 + \beta_1 \mathbf{f_t} + \epsilon_t \, .$$

Statistical Learning Approach

Other ways to achieve sparse models or avoid parameter proliferation: variable/feature/model selection, model averaging, or model ensemble.

- penalized regression: ridge, Lasso (De Mol et al., 2008).
- bagging (Inoue & Kilian, 2008).
- boosting (Bai & Ng, 2007).
- Bayesian model averaging (Koop & Potter, 2004; Wright, 2009).

BSTS-U-MIDAS Model

- State-space component:
    - Bayesian Structural Time Series (BSTS) model decomposes the target variable into several stochastic processes.
- Regression component:
    - U-MIDAS model tackles mixed-frequency data.
    - Spike-and-slab regression is used for variable selection to handle the high dimensionality problem.
    - BMA deals with model uncertainty and instability.

Local Linear Trend Model with Regression

- Observation equation(level + regression):

$$y_t = \mu_t + \beta x_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

- State equation 1 (random walk + trend):

$$\mu_t = \mu_{t-1} + b_{t-1} + w_{1t}, \quad w_{1t} \sim N(0, W_1)$$

- State equation 2 (random walk for trend):

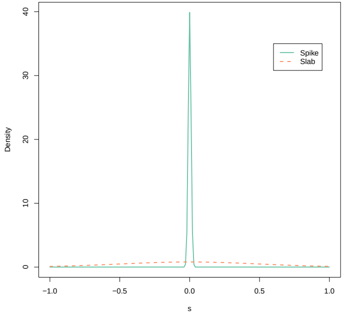
$$b_t = b_{t-1} + w_{2t}, \quad w_{2t} \sim N(0, W_2)$$

BSTS Model cont.

- Assumption:
  - ϵ_t, w_{1t} and w_{2t} are independent,
- States to estimate α:
  - μ_t, b_t; Kalman filter.
- Parameters to estimate θ:
  - θ_1: W_1, W_2 for state component;
  - θ_2: β, σ_ϵ^2 for regression component; spike-and-slab regression.

Spike-and-Slab Prior

$$\beta_i \sim (1-\gamma_i)N(0, c\varphi^2) + \gamma_i N(\tilde{\beta}_i, \varphi^2)$$



$\gamma_i = 0$: irrelevant predictor i has zero β_i. c is very small number. A "spike" at the origin.

$\gamma_i = 1$: relevant predictor i has non-zero β_i. φ^2 is very large. Approximate to a "slab".

$$\gamma_i \sim \pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}.$$

π_i is predictor x_i's probability of inclusion.

Estimation Using a Gibbs Sampler

Alternates between draws of $p(\alpha|\theta, \mathbf{y}_{1:t})$ and $p(\theta|\alpha, \mathbf{y}_{1:t})$.

1. for state α: μ, b.
   - Draw from $p(\alpha|\mathbf{y_{1:t}}, \theta)$ using a stochastic versions of the Kalman smoother from Durbin and Koopman (2002);
2. for parameters θ_1 in state component :
   - Draw for W_1, W_2 from independent inverse Gamma distributions given $\alpha, \mathbf{y_{1:t}}$.
3. for parameters θ_2 in regression component.
   - Draw from $p(\beta, \sigma^{-2}, \gamma) = p(\beta \mid \gamma, \sigma^{-2})p(\sigma^{-2} \mid \gamma)p(\gamma)$ using spike-and-slab prior with stochastic search variable selection(SSVS) algorithm given $\alpha, \mathbf{y_{1:t}}$.

Bayesian Model Averaging

With a sequence of posterior $\{\alpha_1, \theta_1; \alpha_2, \theta_2, ...\}$ from MCMC,

- Get predictive distribution for one step ahead forecast for target variable $p(\tilde{y}_{t+1}|\mathbf{y_{1:t}}, \mathbf{x_{1:(t+1)}})$.
- By averaging over draws for $\tilde{y}_{t+1}$, get a point forecast, which is a form of Bayesian model averaging and accounts for the model uncertainty.
- Average over draws for γ_i to get predictor x_i's probability of inclusion π_i.

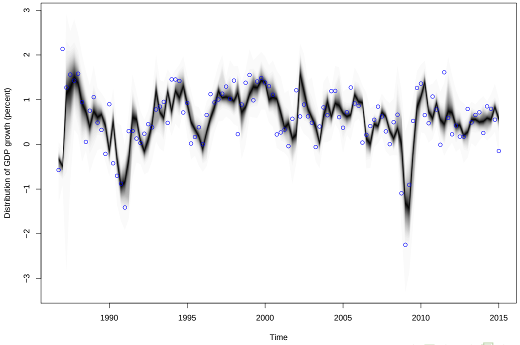
Empirical Application: Data

- Quarterly GDP.
  - Seasonally adjusted.
  - Log differenced for comparison with ARIMA.
- Daily(include $22 \times 12 - 1$ lags):
  - Toronto Stock Exchange (TSX) index,
  - daily West Texas crude oil prices,
- Monthly(include $3 \times 8 - 1$ lags):
  - unemployment rate,
  - monthly spread between interest rates of 10 year government bonds and 3 month treasure bill,
  - monthly housing starts.

Detrend ,deseasonalize, and standardize the predictors.
Total 600 predictors; 114 observations.

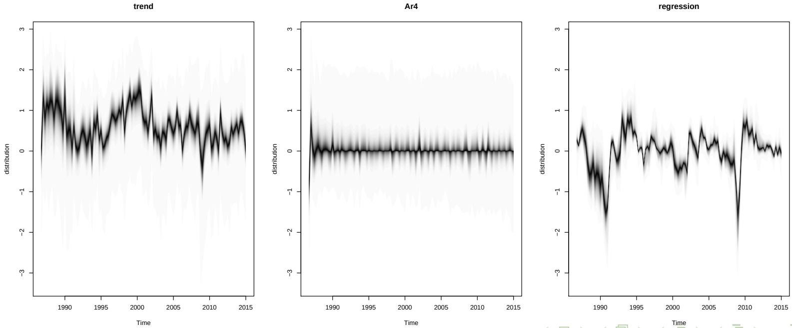
One Step Ahead Forecast of GDP Growth

MCMC: 10,000 iterations and discarded the first 2,000 as burn in.
The default priors in R package "bsts" (Scott, 2015), except choosing the "expected model size" to 4.



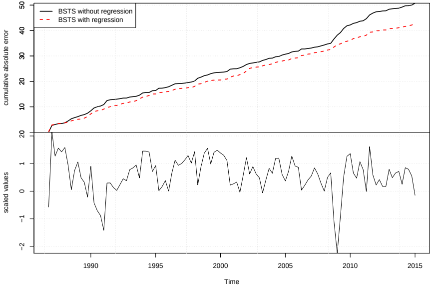
Contributions of States for GDP Growth

Add an AR(4) term in the state equation. The AR term is relatively stable, and the trend component is more volatile. The regression component exhibits the most variation, which helps to capture the turning points of the GDP growth.



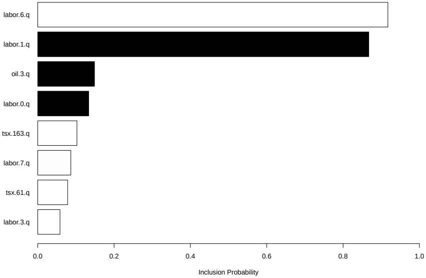
The Cumulative Absolute One Step Ahead Forecast Error

During 2008 to 2009 financial crisis, the cumulative error for model without regression component increases rapidly, however the error for BSTS with regression increases at a constant rate, which shows its robustness.



Predictors with High Inclusion Probability

The inclusion probability of predictor indicates its ability of helping prediction. A white bar indicates positive relationship, black indicates negative.

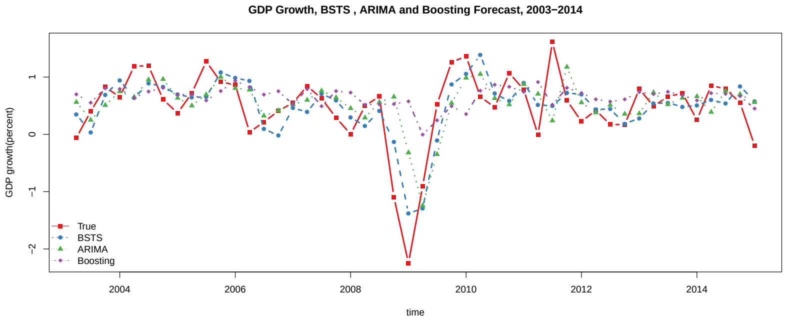


Year	2014						2015				
Month	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
GDP								?			
Labor		**7**	**6**			*3*		*1*	*0*		
TSX						**163**			**61**		
Oil											*3*

It shows most of models are sparse. A combination of high frequency data works as a good predictor, which is similar to MIDAS weighting scheme and factor.

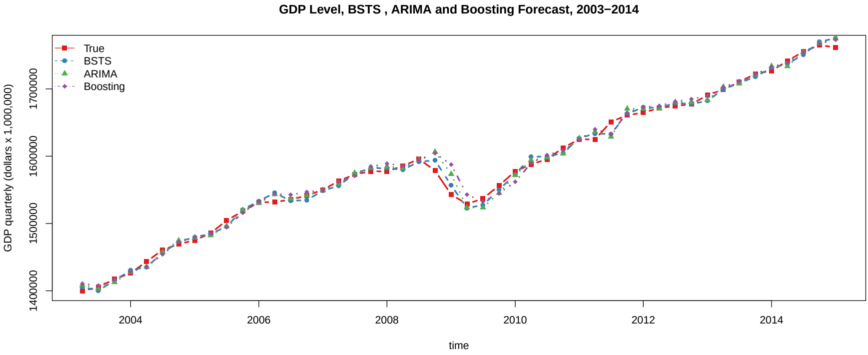
Comparison with ARIMA, and Boosting

Choose 2003 to 2015 as test period. ARIMA is benchmark; Boosting model is flexible; nonlinear .



BSTS outperforms other two due to better performance (MAPE) during 2008 to 2009 financial crisis.

Comparison in GDP Level



	BSTS	ARIMA	Boosting
MAPE	0.356	0.401	0.405

Conclusions

Advantage of BSTS-U-MIDAS:

- flexible for handling high frequency and ragged data,
- capable of capturing the structural breaks or turning points,
- robust to irrelevant or redundant variables,
- easy to incorporate new information of predictors to update the forecast.

Further study:

- test the model based on simulated data,
- use historical data to improve the prior,
- model volatility of economic variable instead of the level or return.