Nowcasting by the BSTS-U-MIDAS Model

Jun Duan

Supervised by: David Giles

September 3, 2015

University of Victoria
Economics

Outline

Forecasting with Mixed Frequency Data

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

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

BSTS-U-MIDAS Model

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

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

Empirical application: forecast quarterly GDP for Canada. Accurate than ARIMA and Boosting model. Capture the structural breaks in the 2008-2009 crisis.

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 Models

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

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

$$\text{Alignment: } \begin{bmatrix} y_{2nd\ quarter} & | & x_{June} & x_{May} & x_{Apr} \\ y_{1st\ quarter} & | & x_{Mar} & x_{Feb} & x_{Jan} \\ \cdots & | & \cdots & \cdots & \cdots \end{bmatrix}$$

- 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).

dynamic factor model

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

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

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

High frequency f_t is aggregated to low frequency $\mathbf{f_t}$. Usually, solved by maximum likelihood estimation with the Expectation Maximization algorithm in a state-space representation.

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

- Bayesian Structural Time Series (BSTS) model captures the dynamic feature of the target variable. The Kalman filter is used to formulate the likelihood function.
- 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.

BSTS Model

Local linear trend model with regression

- Observation equation(level + regression):

$$y_t = \mu_t + z_t + v_t, \quad v_t \sim N(0, V)$$

- 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.

- Regression component: $z_t = \beta x_t$
- Parameters to estimate: $\theta : V, W_1, W_2$ for state component; β, σ^2 for regression component
- States to estimate: $\alpha : \mu_t, b_t$

Optimal Kalman forecast

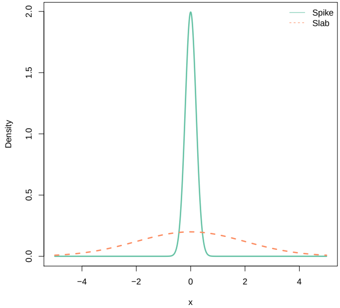
$$\hat{y}_t = \mu_t + z_t + K_t(\hat{y}_{t-1} - y_{t-1})$$

where K is optimal Kalman gain which is a function of the variance terms V_1, W_1, W_2 .
Regression component: $z_t = \beta x_t$ will be taken care of by a spike-and-slab regression.

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 larger. A "slab" of constant height.

$\gamma \sim \prod_i \pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}$. π_i is predictor x_i's probability of inclusion in the regression.

Estimation using a Gibbs sampler

Simulate α, θ from a Markov chain with $p(\alpha, \theta|y)$.

1. for state α such as 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. parameter-simulation for θ such as V, W_1, W_2 for state component and $\gamma, \beta_\gamma, \sigma^2$ for regression component:
   - draws for V, W_1, W_2 from conjugate Gamma distribution given $\alpha, \mathbf{y_{1:t}}$.
   - draw for γ using spike-and-slab prior. Draw for β_γ, σ^2 using a conjugate normal-inverse Gamma distribution with Zellners g-prior given γ.

Bayesian Model Averaging

With a sequence posterior α_i, θ_i from MCMC, we also get posterior distribution of one step ahead forecast for target variable:

$$p(\tilde{y}_{t+1} | \mathbf{y_{1:t}}, \mathbf{x_{1:(t+1)}})$$

for each time point t. By averaging these draws for $\tilde{y}_{t+1}$, we can get a point forecast, which is a form of Bayesian model averaging and accounts for sparsity and model uncertainty.
We also can average over draws for γ_i to get predictor x_i's probability of inclusion in models.

Empirical Application: data

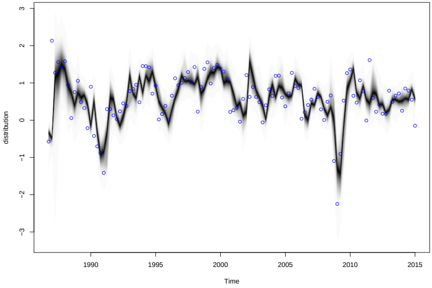
We use 5 different frenquency macroeconomic predictors to forecast quarterly GDP for Canada.

- quarterly GDP for Canada (1981/03 - 2015/01).
- daily Toronto Stock Exchange (TSX) index,
- daily West Texas crude oil prices,
- monthly unemployment rate,
- monthly spread between interest rates of ten years government bonds and three month treasure bill,
- monthly housing starts data.

We detrend ,deseasonlize, and standardize the predictors. For daily data, we include $22 \times 12 - 1$ lags. For monthly data, $3 * 8 - 1$ lags. Total 602 predictors in regression component.

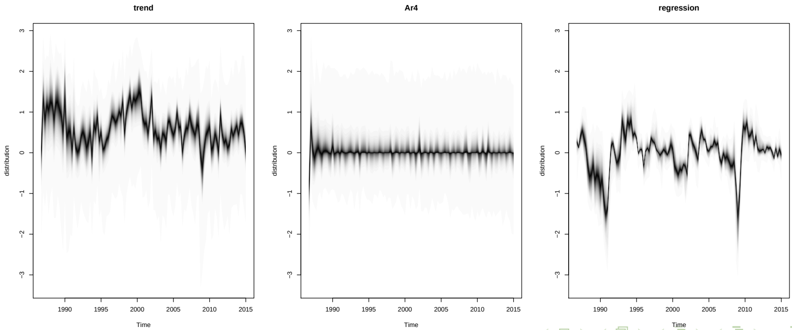
One Step Ahead Forecast

We ran the MCMC algorithm for 10,000 iterations and discarded the first 1,000 as burn in. We used the default priors in R package "bsts" (Scott, 2015), except we choose the "expected model size" to 4. We choose 2003 to 2015 as test periods.



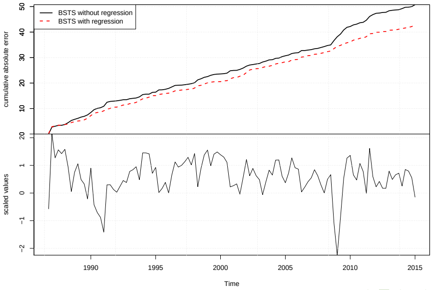
Contributions to State for GDP Growth

We also added a 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 in the dynamics of the GDP growth.



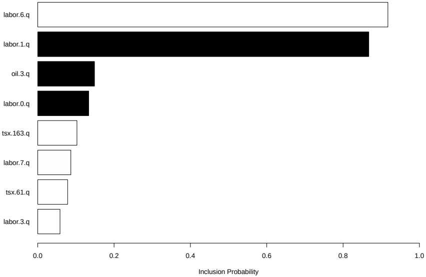
The Cumulative Absolute One Step Ahead Forecast Error

During 2008 to 2009 financial crisis, the cumulative error for model without regression component increases rapidly, however the error for our model increases in 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.

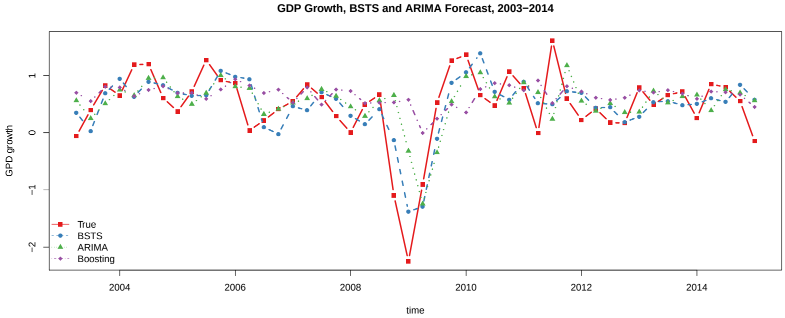


Month	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
TSX					**163**			**61**		
Oil										*3*
Labor	**7**	**6**			**6**		*1*	*0*		
GDP	Quarter 3		Quarter 4			Quarter 1			Quarter 2	
Year	2014					2015				

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

ARIMA is benchmark; Boosting model is more flexible and can capture nonlinear relationship.



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

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 simulation data,
- use historical data to improve the prior,
- model volatility of economic variable instead of the level or return.