High-dimensional Google Queries for Nowcasting New Housing Sales

Application of Bayesian Structural Time-Series Model

MAI-ANH DANG & HOANG TRAN

M2 EEE 2017-2018 | TOULOUSE SCHOOL OF ECONOMICS

1. Motivation

- **Nowcasting & Short-term Forecasting:**
  - The official are usually available with publication lags
  - Desire a method to timely estimate the current values (i.e. New Housing Sales)
- **Google Queries: Potential Predictors**
  - The data is nearly real-time
  - Potential queries contemporaneously correlated with our interest series
  - ⇨ *Useful to nowcast the demand of new houses (Choi & Varian 2009, 2012)*
- **High-dimensional Issues in Time-series context**
  - Google Correlate enables us to derive the hundred of most correlated
  - Spurious correlated terms (expect the sparsity + variable selection)
  - Time-series data: Serial Correlation

2. Bayesian Structural Time Series

Observation	$y_t = \mu_t + \beta^T x_t + \varepsilon_t$	$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$
Regression component	$\beta^T x_t$	
Trend + Random Walk	$\mu_t = \mu_{t-1} + \delta_{t-1} + u_t$	$u_t \sim N(0, \sigma_u^2)$
Random Walk	$\delta_t = \delta_{t-1} + \nu_t$	$\vartheta_t \sim N(0, \sigma_\nu^2)$

- Easy to add the Seasonality component:

$$\tau_t = -\sum_{s=1}^{s-1} \tau_{t-s} + w_t \,, where\ w_t \sim N(0, \sigma_w^2)$$

- The BSTS method (attempt to estimate the posterior probability of models)
  - Structural Time-Series model for target series
  - Spike-and-Slab Regression (estimate the inclusion prob. of each variable)
  - Markov Chain Monte Carlo Simulation

2.0. State-space form

Observed (1)	$y_t = Z_t^T \alpha_t + \varepsilon_t$	$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$
y_t	Z_t^T	α_t
	$(1 \quad 0 \quad \beta^T x_t)$	$(\mu_t \quad \delta_t \quad 1)'$
Transition (2)	$\alpha_t = T_t \alpha_{t-1} + N_t \eta_t$	$\eta_t \sim N(0, Q_t)$
α_t	$T_t \alpha_{t-1}$	$N_t \eta_t$
$\begin{pmatrix} \mu_t \\ \delta_t \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \mu_{t-1} \\ \delta_{t-1} \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} u_t \\ \nu_t \\ 0 \end{pmatrix}$

- It is assumed that the time development of target series depends on unobserved state α
- The unobserved state α_t would be obtained by **Kalman approach**
- y_t^* (after subtracting time components), conduct **the Spike-and-Slab**

2.1. Structural Time Series

Kalman Filter Recursion

- Move forward from t=1,…,T
- Compute predicted:

$$p(\alpha_{t+1}|y_{1:t})$$

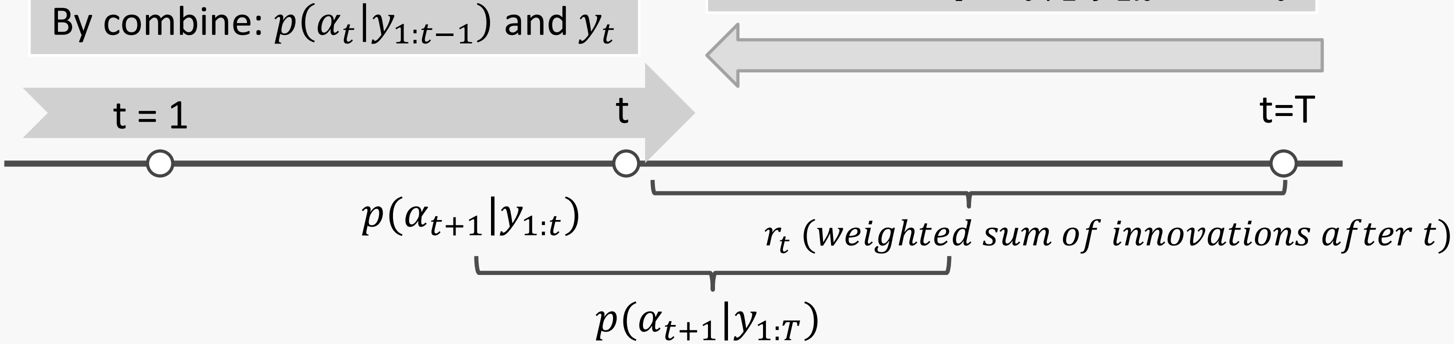
By combine: $p(\alpha_t|y_{1:t-1})$ and y_t

Kalman Smoother

- Move backward from t=T,…,1
- Compute:

$$p(\alpha_{t+1}|y_{1:T})$$

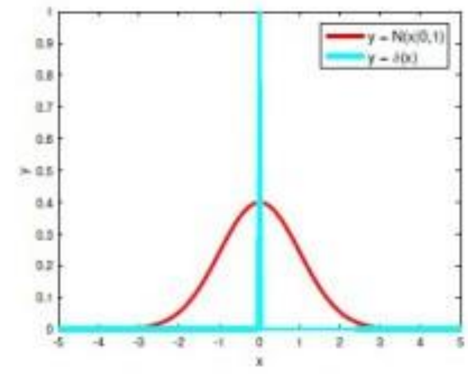
By combine: $p(\alpha_{t+1}|y_{1:t})$ and r_t



- Durbin & Koopman (2002) algorithm enables simulating α_t from $p(\alpha_t|y)$, taking into account the serial correlation
- We can obtain the posterior distribution $p\left(\frac{1}{\sigma_u^2}, \frac{1}{\sigma_v^2} \mid \alpha, y\right)$ (Scott & Varian 2014)

2.2. Spike-and-Slab Regression

- Let $\gamma_k = 1, if\ \beta_k \neq 0, \gamma_k = 0\ otherwise$
- β_γ denote the subset of elements β where $\beta_k \neq 0$
- **Joint spike-and-slab prior distribution:**



(b) Spike and slab prior

$$p(\beta, \gamma, \sigma_\varepsilon^2) = p\left(\beta_\gamma \middle| \gamma, \sigma_\varepsilon^2\right) p(\sigma_\varepsilon^2 \mid \gamma)\, p(\gamma)$$

Prior Distribution	Posterior Distribution
$p(\gamma) = \prod_{k=1}^{K} \pi^{\gamma k}(1-\pi)^{1-\gamma k}$	$\gamma \mid \boldsymbol{y}^*$ (obtained by **analytical** marginalize over β_γ and $\frac{1}{\sigma_\varepsilon^2}$)
$\frac{1}{\sigma_\varepsilon^2} \mid \gamma \sim Ga\left(\frac{df}{2}, \frac{ss}{2}\right);\ \frac{ss}{df} = (1-R^2)s_y^2$	$\frac{1}{\sigma_\varepsilon^2} \mid \gamma, \boldsymbol{y}^* \sim Ga\left(\frac{DF}{2}, \frac{SS_\gamma}{2}\right)$
$\beta_\gamma \mid \gamma, \sigma_\varepsilon^2 \sim N\left(b_\gamma, \sigma_\varepsilon^2 \left(\Omega_\gamma^{-1}\right)^{-1}\right);\ \Omega^{-1} \propto X'X$	$\beta_\gamma \mid \gamma, \sigma_\varepsilon^2, \boldsymbol{y}^* \sim N\left(\widetilde{b_\gamma}, \sigma_\varepsilon^2 \left(V_\gamma^{-1}\right)^{-1}\right)$

2.2. Spike-and-Slab Regression

Posterior Marginal Distribution $p(\gamma)$:

$$\gamma \mid \boldsymbol{y}^* \sim \mathrm{C}(\boldsymbol{y}^*) \frac{|\Omega_\gamma^{-1}|^{\frac{1}{2}} p(\gamma)}{|V_\gamma^{-1}|^{\frac{1}{2}} SS_\gamma^{\frac{DF}{2}-1}}$$

- $\mathrm{C}(\boldsymbol{y}^*)$: normalizing constant depending on $\boldsymbol{y}^*$
- V_γ^{-1}: low dimensional (if the model is sparse)
- Different from $\mathcal{L}_1$-regularization in LASSO, the variable selection mechanism happens as we place a positive probability on coefficients being zero

⇨ ***The Sparsity is featured by the full posterior distribution, not simply by the setting value***

- Posterior Distribution cannot compute directly (the sum over the model space is intractable)

⇨ ***Using the MCMC***

2.3. Markov Chain Mote Carlo Simulation

- MCMC is used to obtain the posterior distribution (which is difficult to analytically obtain)

$$\theta\text{: } sets\ of\ parameters\ (\beta, \gamma, \sigma_\varepsilon^2)$$
$$p(\theta|y) \propto p(y|\theta)p(y)$$
$$posterior\ distribution \propto likelihood\ \times prior\ distribution$$

- MCMC is the idea to randomly sample under a special **sequential process**. The algorithm is designed in the manner that the next random sample will depend on the previous random sample (as a **chain**)

- **Intuition:** MCMC algorithm “walks” through the model space, models with higher posterior distribution will be visited more often

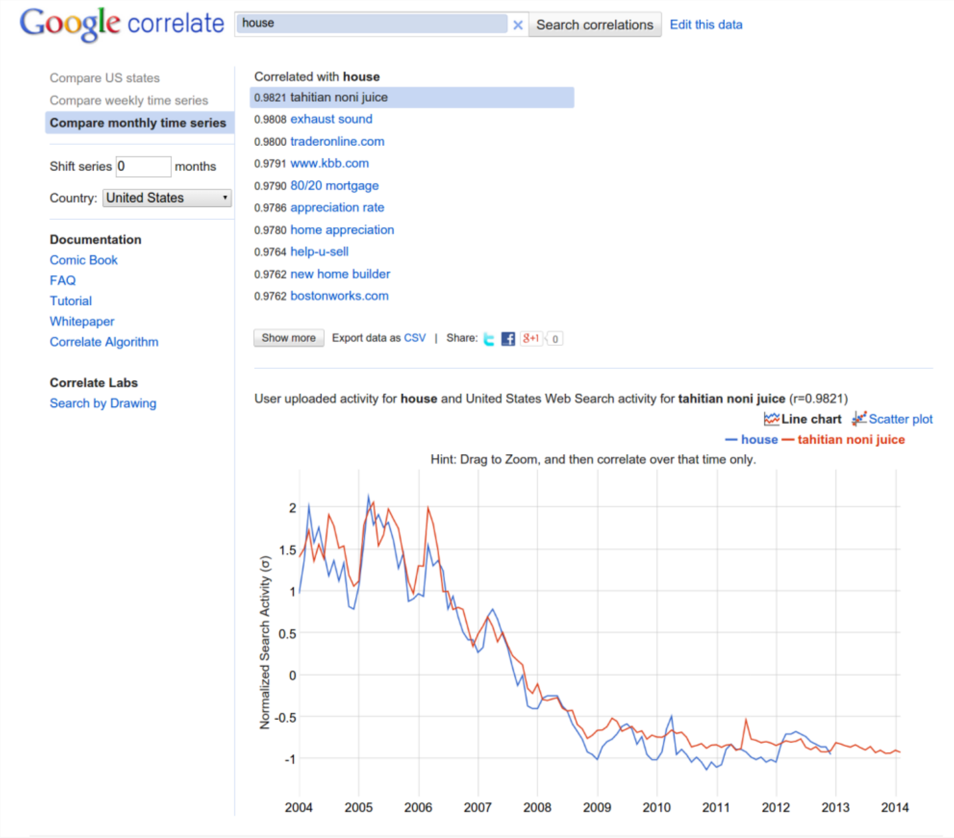
- **Variable Selection Mechanism:** the prior inclusive probability for each predictor is updated in each step of MCMC, more important variable would appear more often

2.3. Markov Chain Mote Carlo Simulation

- The posterior distribution can be simulated by the MCMC Algorithm, following this step:
  - Starting point: simulate $\gamma, \beta, \sigma_\varepsilon^2, \sigma_\upsilon^2, \sigma_u^2$ from prior distribution
    1. Simulate α from $p(\alpha | y, \gamma, \beta, \sigma_\varepsilon^2, \sigma_\upsilon^2, \sigma_u^2)$ using Durbin & Koopman(2002)
    2. Simulate σ_u^2 and σ_ν^2 from their posterior distribution $p\left(\frac{1}{\sigma_u^2}, \frac{1}{\sigma_\nu^2} \mid y, \alpha, \beta, \sigma_\varepsilon^2\right)$
    3. Simulate β and σ_ε^2 from their posterior distribution $p(\beta, \sigma_\varepsilon^2 | \, y, \alpha, \sigma_u^2, \sigma_\nu^2)$
  - With updated $\gamma, \beta, \sigma_\varepsilon^2, \sigma_\upsilon^2, \sigma_u^2$: Back to step 1
- For each step, we obtain $\phi = (\gamma, \beta, \sigma_\varepsilon^2, \sigma_\upsilon^2, \sigma_u^2)$. By M steps of MCMC, we obtain $\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(M)}$ from a Markov chain with stationary distribution $p(\phi | y)$
- By each set of parameters (ϕ), obtain the forecast value $\hat{y}$
- Averaging all over the models for the final predictions

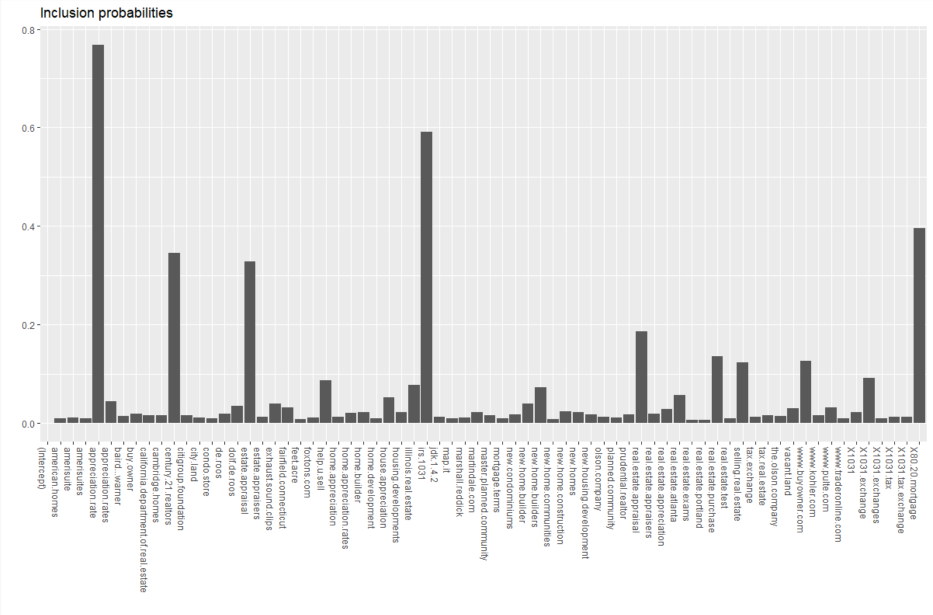
3. Data

- The *"New One Family House Sold"* data will be downloaded from FRED, for the period 01-01-04 to 01-09-12
- The series will then be feed to Google Correlate
- **100 most correlated queries** to the Housing Price will be returned
- Some spurious queries will be removed manually (e.g. "tahitian noni juice", "exhaust sound",...)
- End up with **70 predictors**



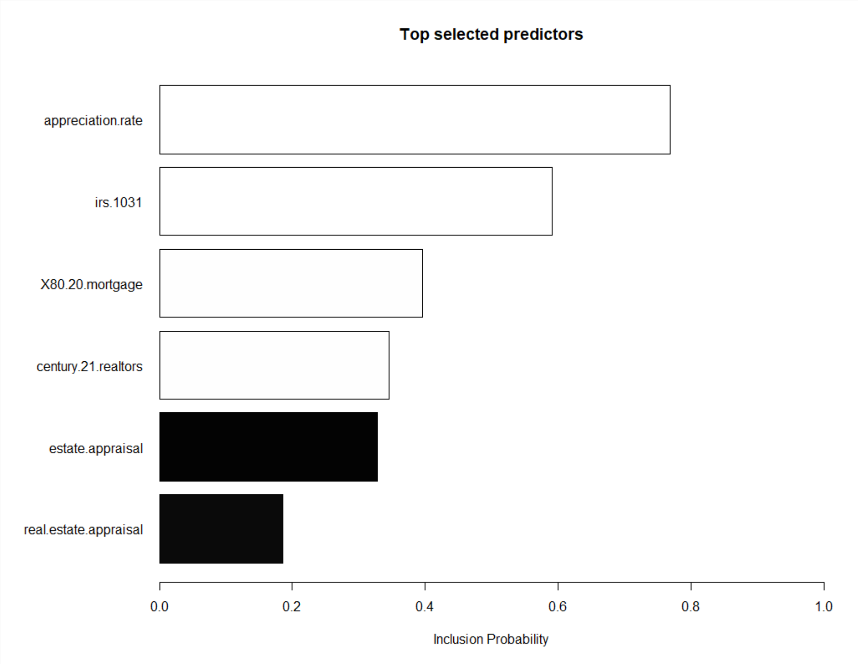
4.1 Variables Selection

- Top predictors selected based on Inclusion Probabilities
- Predictors with high Inclusion Probabilities are more important
- Sparsity of the model
- Will be used to compare with other variable selection like Ridge, LASSO, Elastic Net



4.1 Variables Selection

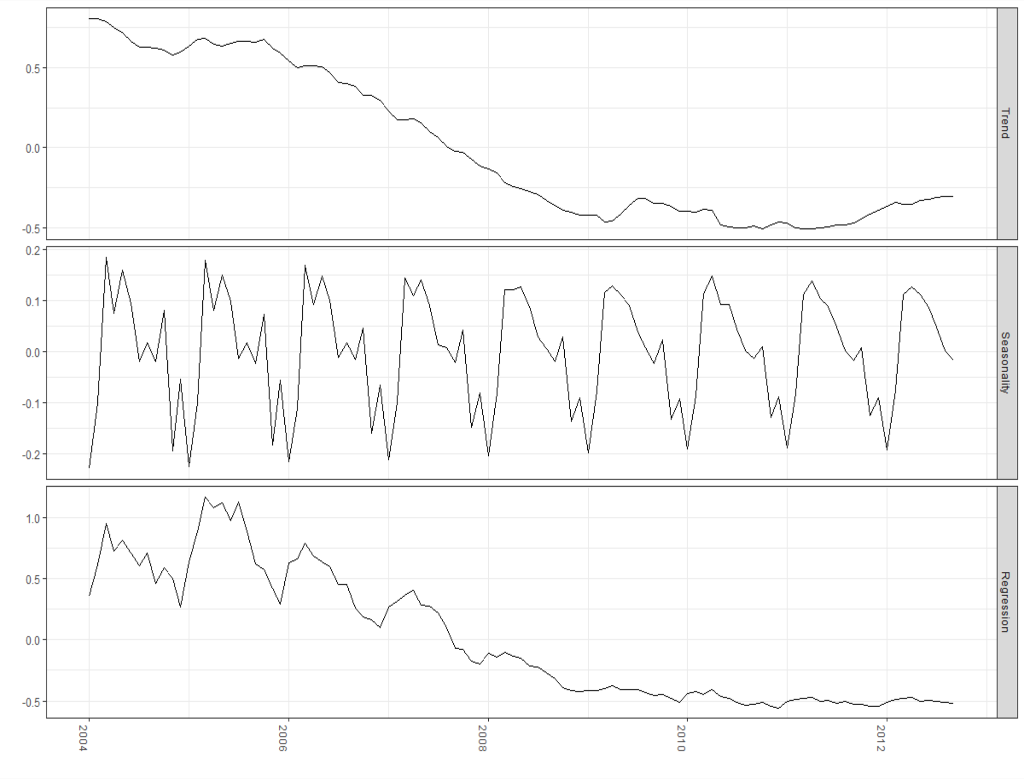
- Top predictors selected based on Inclusion Probabilities
- Predictors with high Inclusion Probabilities are more important
- Sparsity of the model
- Will be used to compare with other variable selection like Ridge, LASSO, Elastic Net



4.2 Components Contribution

One clear advantage of using BSTS over ARIMA is **its ability to derive the contribution of each components to the model**.

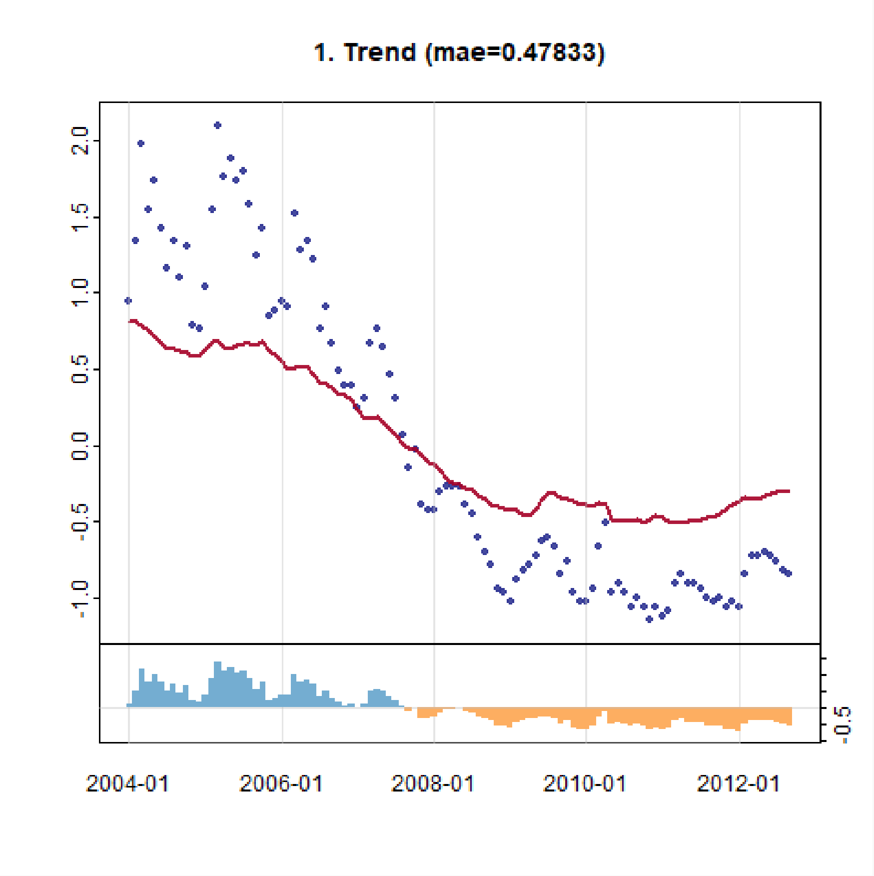
- Downward Trend
- Obvious seasonality pattern
- Regression predictors contribute significantly to the model



4.2 Components Contribution

Each component will be combined (stacked) in order to derive the final estimation

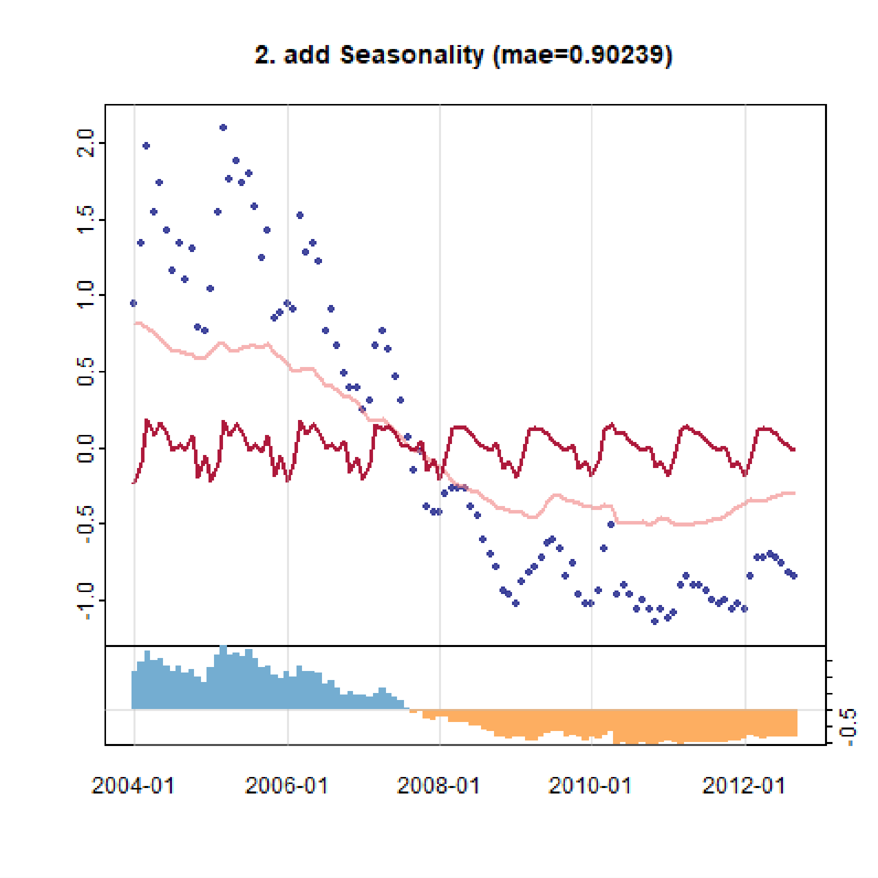
First, the **Trend component**



4.2 Components Contribution

Each component will be combined (stacked) in order to derive the final estimation

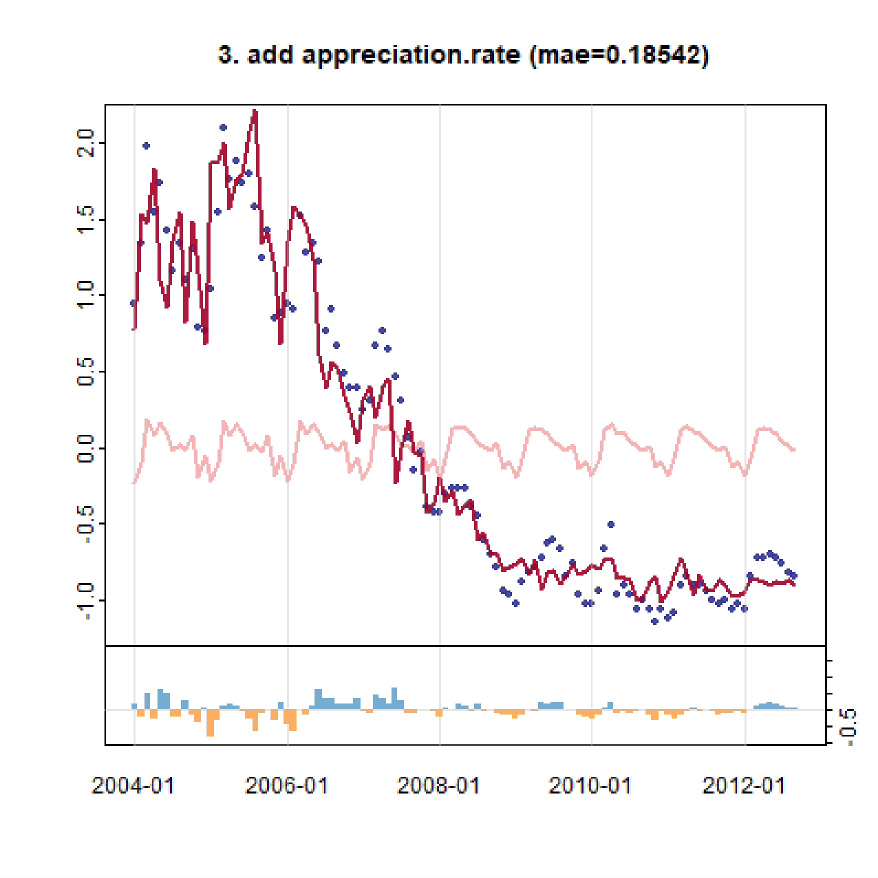
Then, the **Seasonality component** will be added to the model



4.2 Components Contribution

Each component will be combined (stacked) in order to derive the final estimation

After that, our **top predictor** “appreciation.rate” will be included

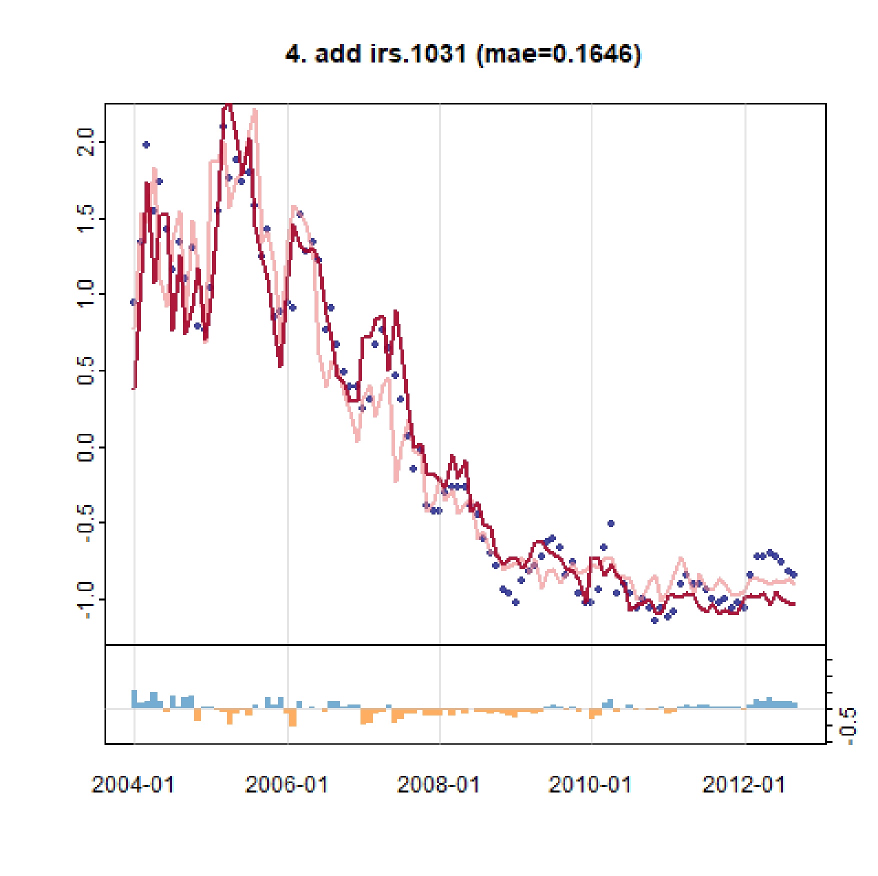


4.2 Components Contribution

Each component will be combined (stacked) in order to derive the final estimation

Finally, our **top predictor** "irs.1031" will be added

=> With only 2 predictors included, we already have good fit



5. Out-of-sample: AR and BSTS

Consider 2 models:

- Baseline AR using lag 1 and 12:

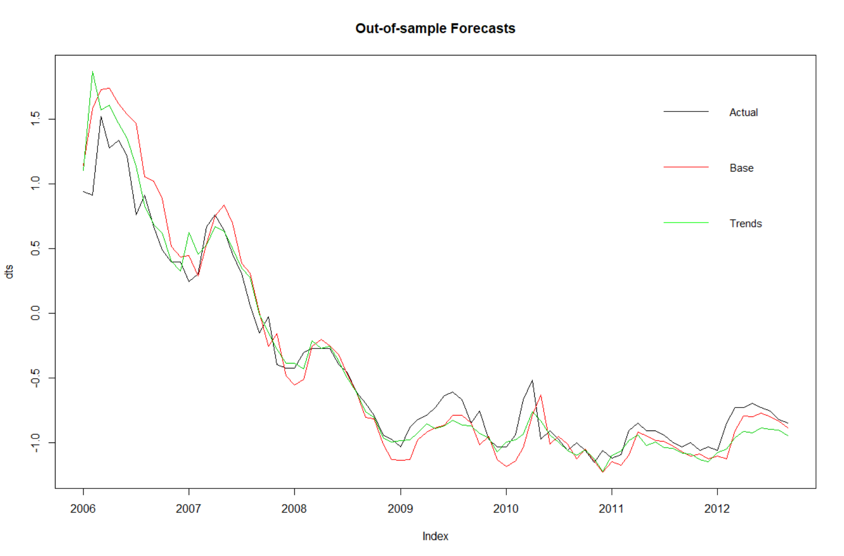
$$y_t = b_1 y_{t-1} + b_{12}\, y_{t-12} + e_t$$

- Same model but adding some top predictors from Google Correlate:

$$y_t = b_1 y_{t-1} + b_{12}\, y_{t-12} + a_t x_t + e_t$$

- Mean Absolute Percent Error (MAPE) for each model is utilized for comparision.

=> Model using Google predictors derive significantly lower prediction error.



mae.base	mae.trends	mae.delta
0.1451080	0.1115476	0.2312789

6. Comparison with other regularization methods

Predictors	BSTS	Ridge	LASSO	Elastic Net
appreciation.rate	0.768	1	1	1
irs.1031	0.591	7	3	3
X80.20.mortgage	0.395	3	4	4
century.21.realtors	0.345	19	-	-
estate.appraisal	0.327	67	-	-

- Relatively similar selected predictors between 4 systems
- BSTS provides more reliable results, as the other methods base on naïve assumption: treating the time-series data as cross-sectional and ignore trending and seasonality factors

7. Conclusions

PROS

- ❖ Designed to solve variable selection with time-series data
- ❖ All methods in the system have natural Bayesian interpretations and tend to play well together
- ❖ Give better out- of- sample forecasting performance than using a single complex model
- ❖ Superior in dealing with high-uncertainty model

CONS

- ❖ Debatable number of MCMC to converge to stationarity
- ❖ Independent simulated MCMC could create different outcomes
- ❖ There might be some residual effect of starting position / prior distribution
- ❖ Spike-and-slab model produces inclusion probabilities, which might cause difficulties in analysis and comparison

References

- Hastie, T., & Qian, J. (2014). *Glmnet Vignette.* Retrieved from http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html
- Scott, S. L., & Varian, H. R. (2013). Predicting the Present with Bayesian Structural Time Series.
- Scott, S. L., & Varian, H. R. (2014). Bayesian Variable Selection for Nowcasting Economic.
- Stephens-Davidowitz, S., & Varian, H. R. (2015). A Hands-on Guide to Google Data.
- Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 3-28.
- Larsen, K. (2016). *Sorry ARIMA, but I'm Going Bayesian*. Retrieved from https://multithreaded.stitchfix.com/blog/2016/04/21/forget-arima/#footnote2