Hierarchical Model with Bayesian Method for Sales Growth Rate prediction

Author: Diheng Chen

dc3829

Target

Alternative data such as credit card and web data allow investors to monitor company fundamentals on a real-time basis. So, a model has good performance on forecasting a company's actual reported sales growth is very useful and meaningful. Here, we want to use panel sales data of multiple companies to make prediction on a company's sales growth rate on next delivery date.

Problem Analysis

The example of data we can use are shown below.

delivery_date	tab	label_ordered	sales_actual_yoy	panel_sales
12/16/2015	A	15Q4	0.03	-0.05
12/30/2015	A	15Q4	0.03	-0.02
4/20/2016	A	16Q1	0.07	0.03
5/4/2016	A	16Q1	0.07	0.01
6/15/2016	A	16Q2	0.03	-0.03
6/29/2016	A	16Q2	0.03	-0.02
7/6/2016	A	16Q2	0.03	-0.006
12/30/2015	D	15Q4	0.03	-0.05
1/6/2016	D	15Q4	0.03	-0.05
5/4/2016	D	16Q1	0.06	0.009
5/18/2016	D	16Q1	0.06	0.01
6/15/2016	D	16Q2	0.07	-0.02
6/29/2016	D	16Q2	0.07	-0.01
7/6/2016	D	16Q2	0.07	0.008
9/21/2016	L	16Q3	0.09	0.01
2/2/2017	L	16Q4	0.19	0.04
2/16/2017	L	16Q4	0.19	0.03
3/16/2017	L	17Q1	0.10	0.03
……	……	……	……	……

The data are modified due to confidentiality.

As shown in the data sample above, we have different companies panel sales data on certain delivery date, but not all companies have data on every delivery date. The actual sales growth rate data is reported every quarter. On a certain date, we can use all companies' panel sales data to forecast a specific company's actual reported sales growth on next delivery date. Thus, the model we used for prediction should be able to distinguish and utilize data from different groups(companies) to predict data for a specific group(company). We found that Hierarchical Model has this characteristic.

Method-Bayesian Method for Multilevel Modeling

Hierarchical or multilevel modeling is a generalization of regression modeling. Multilevel models are regression models in which the constituent model parameters are given probability models.[1] This implies that model parameters are allowed to vary by group, which means we can keep the independence of each company while making use of all companies' data.

Generally speaking, a company's sales growth rate is relatively stable in a certain period of time, which means that sales growth rate has a high correlation with the company's previous performance. So, I add group-level predictors to model this characteristic.

Also, our models are implemented with Bayesian Method to improve the performance of prediction.

Based on the above analysis, I create six models. They all belong to Hierarchical model with Bayesian Method, but different model allows different parameters to vary by group. The formulas of six models are shown below.

Model 1- Hierarchical Varying Intercept Model:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Where, j represents different companies, i represents forecasting date. Using the panel_sales as input variable.

Model 2- Hierarchical Varying Intercept Model with group-level predictor Model

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i$$

$$\alpha_j = \gamma_0 + \gamma_1 benchmark_j + \xi_j$$

$$\xi_j \sim N(0, \sigma_\alpha^2)$$

Using the panel_sales as input variable and the insample means of sales_actual_yoy of each company as benchmarks.

Model 3- Hierarchical Varying Slope Model:

$$y_i = \alpha + \beta_{j[i]} x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Using the panel_sales as input variable.

Model 4- Hierarchical Varying Slope Model with group-level predictor Model

$$y_i = \alpha + \beta_{j[i]} x_i + \epsilon_i$$

$$\beta_j = \gamma_0 + \gamma_1 benchmark_j + \xi_j$$

$$\xi_j \sim N(0, \sigma_\beta^2)$$

Using the panel_sales as input variable and the insample means of sales_actual_yoy of each company as benchmarks.

Model 5- Hierarchical Varying Intercept and Slope Model:

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Using the panel_sales as input variable.

Model 6- Hierarchical Varying Intercept and Slope Model with group-level predictor Model

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i$$

$$\alpha_j = \gamma_0 + \gamma_1 benchmark_j + \xi_{1j}$$

$$\beta_j = \gamma_2 + \gamma_3 benchmark_j + \xi_{2j}$$

$$\xi_{1j} \sim N(0, \sigma_\alpha^2)$$

$$\xi_{2j} \sim N(0, \sigma_\beta^2)$$

Using the panel_sales as input variable and the insample means of sales_actual_yoy of each company as benchmarks.

Three different estimators

1.MAP – Maximum a posteriori

It is a point estimator. It will find a best group of parameters according to their prior distribution as result. In Bayesian statistics, a maximum a posteriori probability (MAP) estimate is an estimate of an unknown quantity, that equals the mode of the posterior distribution. The MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. It is closely related to the method of maximum likelihood (ML) estimation, but employs an augmented optimization objective which incorporates a prior distribution (that quantifies the additional information available through prior knowledge of a related event) over the quantity one wants to estimate. MAP estimation can therefore be seen as a regularization of ML estimation. [2]

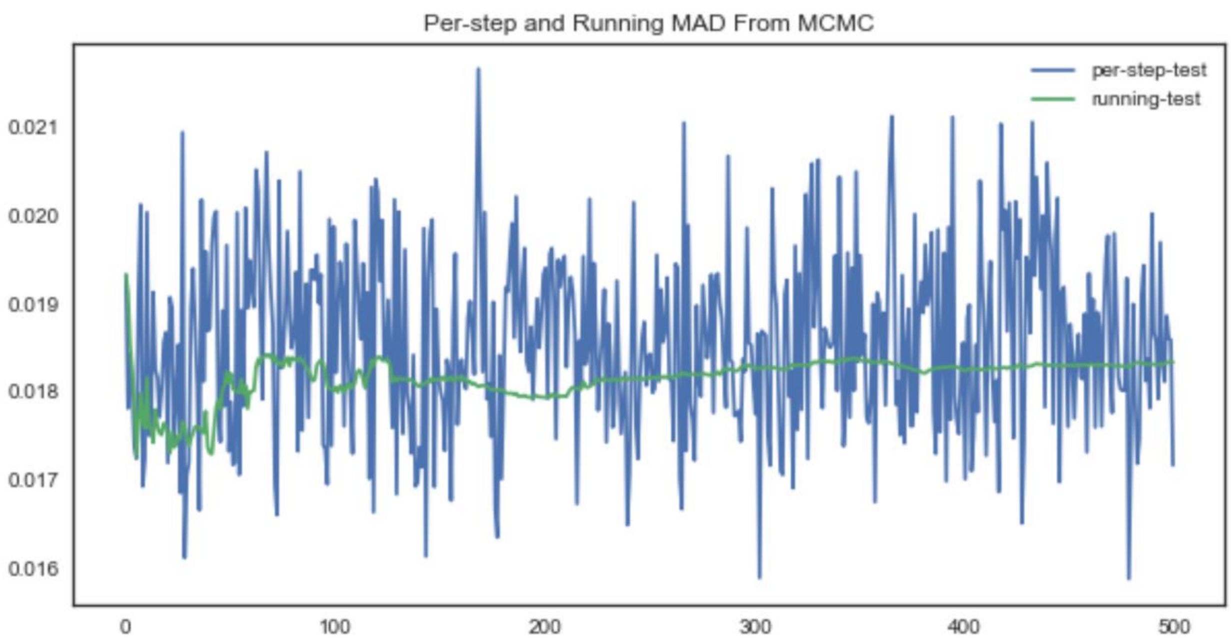
2.MCMC- Markov chain Monte Carlo

In statistics, Markov chain Monte Carlo (MCMC) methods comprise a class of algorithms for sampling from a probability distribution. By constructing a chain that has the desired distribution as its equilibrium distribution, one can obtain a sample of the desired distribution by observing the chain after a number of steps. The more steps there are, the more closely the distribution of the sample matches the actual desired distribution. [3]

It will first draw a lots samples of the possible values of the parameters according to their prior distribution. And then combine the samples to obtain the prediction result. After converge test, we find that 500 samples with 500 tuned samples is enough for Model 1,2,3,4 and 5 to converge. 1000 samples with 500 tuned samples is enough for Model 6 to converge.

Example:

Convergence test picture for Model1:



From the picture, we can see that after 500 samples, the prediction result is converged.

3.ADVI- Automatic differentiation variational inference

The idea is to first automatically transform the inference problem into a common space and then to solve the variational optimization. Solving the problem in this common space solves variational inference for all models in a large class. [4]

It will first fit a distribution for each parameter, and then draw samples according to this distribution. Then, combining the samples to get the prediction result.

Two Approach for Time Series Next Value Predictions

1.Fixed Rolling Window Prediction

Initial train sample contains 6 quarters(15Q4-17Q1), and rolling period is 7 days. Some company may have no data in the next period, therefore, when we are rolling, we must first determine whether a company has data in the next period. The company with data in the next period will adopt the FIFO method, and the company without data remains unchanged for train sample

2. Recursive Rolling Window Prediction

Initial train sample contains 6 quarters(15Q4-17Q1), and rolling period is 7 days. This time the initial data is fixed, and additional observations are added one at a time to the train sample.

One step forward prediction result of six models

Metric: Using Median Absolute Error as metric.

Mean of insample error:

	MAP(Fixed_Roll)	MAP(Recursive)	MCMC(Fixed_Roll)	MCMC(Recursive)
main	0.015660	0.015660	0.015660	0.015660
ss	0.012963	0.012963	0.012963	0.012963
model1	0.010522	0.011918	0.010523	0.011926
model2	0.010513	0.012019	0.012019	0.010500
model3	0.013890	0.014863	0.013814	0.014852
model4	0.013918	0.014998	0.014982	0.013860
model5	0.009083	0.009812	0.008093	0.009697
model6	0.009304	0.010939	0.009800	0.008073

Outsample error：

	MAP(Fixed_Roll)	MAP(Recursive)	MCMC(Fixed_Roll)	MCMC(Recursive)
main	0.01739046	0.01739046	0.01739046	0.01739046
ss	0.01601296	0.01601296	0.01601296	0.01601296
model1	0.01325268	0.01527020	0.01331946	0.01537553
model2	0.01322811	0.01545993	0.01538828	0.01313409
model3	0.01783044	0.02012276	0.01799945	0.02013942
model4	0.01776548	0.02013245	0.02006977	0.01775583
model5	0.01167963	0.01462006	0.01180765	0.01469320
model6	0.01225775	0.01559670	0.01487334	0.01168951

Result Analysis

The best model among all six models using three estimators mention above are Model 5- Hierarchical Varying Intercept and Slope Model using MAP with fixed rolling window prediction and Model 6- Hierarchical Varying Intercept and Slope Model with group-level predictor Model using MCMC with recursive rolling window prediction. Their MAE are 0.01168 and 0.1169 relatively respectively, which are better than the benchmark 0.01691.

Reference:

[1] Peadar Coyle, "A Hierarchical model for Rugby prediction", 29 Dec. 2017.

[2] Maximum a posteriori estimation, https://en.wikipedia.org/wiki/Maximum_a_posteriori_estimation

[3] Markov chain Monte Carlo, https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo

[4] Alp Kucukelbir, Automatic Differentiation Variational Inference, 2 Mar 2016