Final Project Edward Part

Edward

4/20/2024

Group Member Duty

Edward: Literature Review Limitation, and summary parts for the project. Comparing the MCMC model with traditional time series model to predict the stock market in Canada.

Literature Review

For our project on predicting stock prices, we're drawing on the research by Chérief-Abdellatif and Alquier from 2018. This paper is really helpful because it tells us that Variational Bayes (VB) methods are solid, even when we've got a ton of data to work through or the models get complicated. VB is especially feasible when the usual way of doing things such as Markov Chain Monte Carlo (MCMC) is stuck by too much data or too many details in the model.

In addition, the Organisation for Economic Co-operation and Development (OECD) also tracks many sorts of economic stuff. We've got market data from places like Yahoo Finance. This means we can see how changes in the economy, like jobs or production, might influence stock market prices, making our predictions more reliable. Combining these tools with the latest data helps us understand what's really driving changes in the stock market trends.

In short, we're blending the statistical methods we covered in this course and some economic data to tackle the tricky task of guessing where stock prices might go. This mix helps us build a strong, data-driven approach to forecasting that makes sense of complex market movements.

Markov Chain Monte Carlo (MCMC)

After research in Variational Bayes (VB), the following research will show how to forecast the stock market by applying the Markov Chain Monte Carlo (MCMC) model. The methodology includes constructing a SARIMA model as a "control", and then an MCMC model will clearly show if there are any disparities between those two models.

This piece of analysis begins by examining the performance of the SARIMA model, followed by an exploration of its residuals to ensure no underlying patterns are missed.

1. The first plot is the trend by applying the SARIMA model:

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## Series: tsx_series
## ARIMA(0,1,0)(2,0,0)[12] with drift
##
## Coefficients:
##          sar1     sar2    drift
##       -0.0972  -0.0961  40.9611
```

```
## s.e.   0.0470   0.0493  16.9616
##
## sigma^2 = 191779:  log likelihood = -3531.62
## AIC=7071.24   AICc=7071.32   BIC=7087.86
##
## Training set error measures:
##                      ME      RMSE      MAE       MPE     MAPE      MASE
## Training set -0.1453855 436.0664 287.4322 -0.2742959 3.049632 0.2418806
##                    ACF1
## Training set 0.01632222
```

Forecasts from ARIMA(0,1,0)(2,0,0)[12] with drift
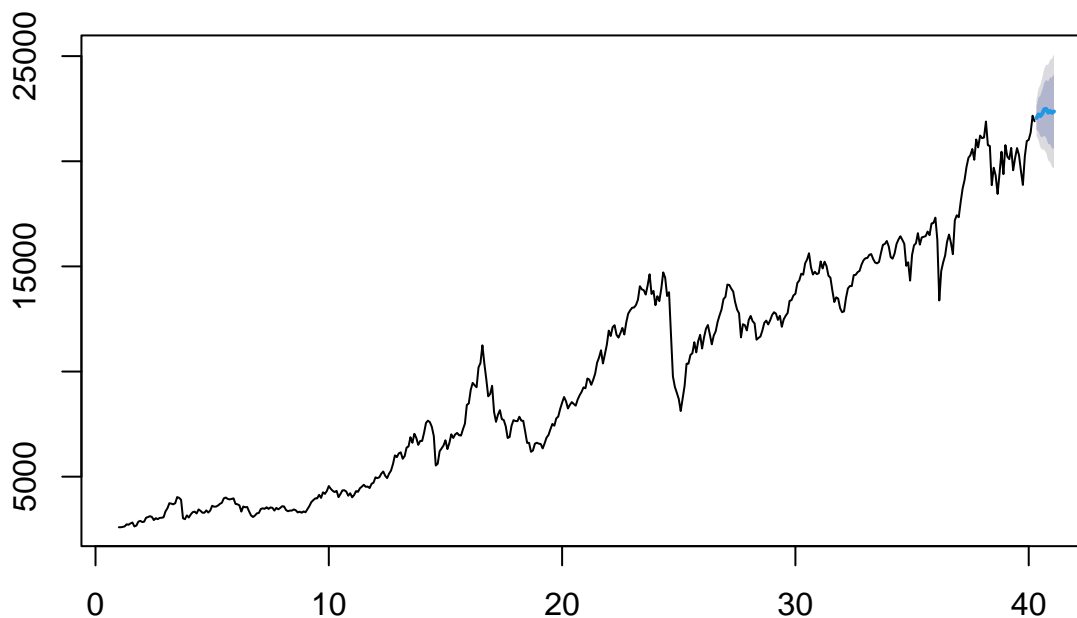
The first plot shows the performance of the SARIMA model. It predicts the trends of the S&P/TSX Composite Index. We can see the seasonal and non-seasonal effects from the plot now. By adjusting the parameters p,d,q for the non-seasonal component and P, D, Q for the seasonal component, we can detect a very obvious trend in this graph. It's an upward-sloping trend. The shadowed area towards the end of the plot likely represents the forecasted values. In general, the stock market goes up even though many small fluctuations occur. We are uncertain about if there is a seasonal effect by looking at this plot itself, however. This model effectively handles the cyclic changes in the market, making it suitable for short-term forecasting where these patterns are prominent.

2. The second plot is the residuals, ACF, and PACF for the SARIMA model:

residuals(sarima_model)
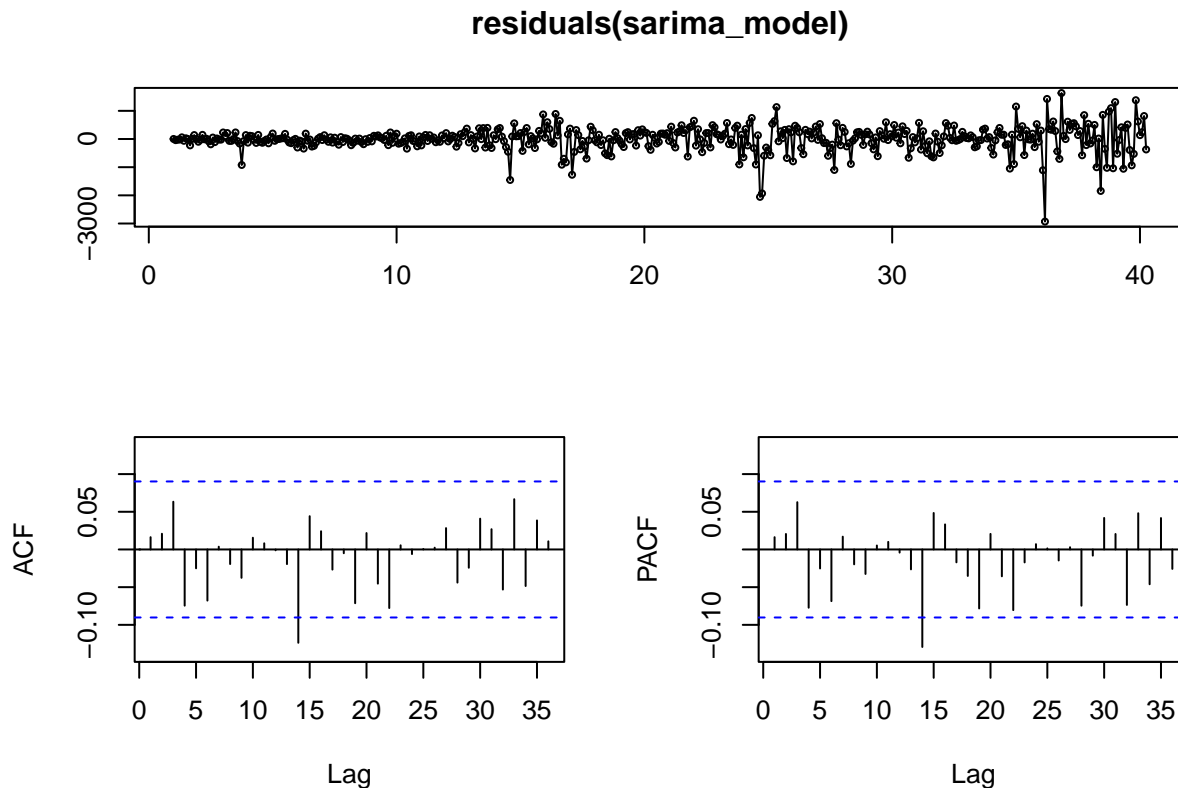
The residuals appear to be centered around zero and do not display any obvious patterns or trends, which means the SARIMA model above captures the trend and fits the data decently. Ideally, the residuals of a well-fitted time series model should resemble white noise and not be biased.

The ACF plot is used to identify any autocorrelation in the residuals at different lags. In this case, the majority of the autocorrelations are within the confidence bounds (dotted blue lines), which implies that there is no significant autocorrelation at various lags. Only one lag spikes out but they are within the bounds in the big picture.

The PACF plot provides a very similar look, the PACF bars are mostly within the confidence intervals, indicating no significant partial autocorrelation at the lags presented.

The residual plot and ACF are used to verify this assumption. A random pattern in the residuals and no significant peaks in the ACF plot beyond the confidence intervals suggest that the model residuals do not have autocorrelation. This plot confirms the model fits my dataset well.

But still, since there is no lags spiking out in the ACF or PACF plots, we are not able to determine the seasonal effect. Recall that the seasonality could be detected by looking at the lags spiking out.

## 3. The third plot is the MCMC model to run a comparison between actual and predicted:

```
## Loading required package: rjags
```

```
## Loading required package: coda
```

```
## Linked to JAGS 4.3.2
```

```
## Loaded modules: basemod,bugs
```

```
## Compiling model graph
```

3

```
##     Resolving undeclared variables
##     Allocating nodes
## Graph information:
##     Observed stochastic nodes: 377
##     Unobserved stochastic nodes: 4
##     Total graph size: 2273
##
## Initializing model
```

Actual vs Predicted Close Prices
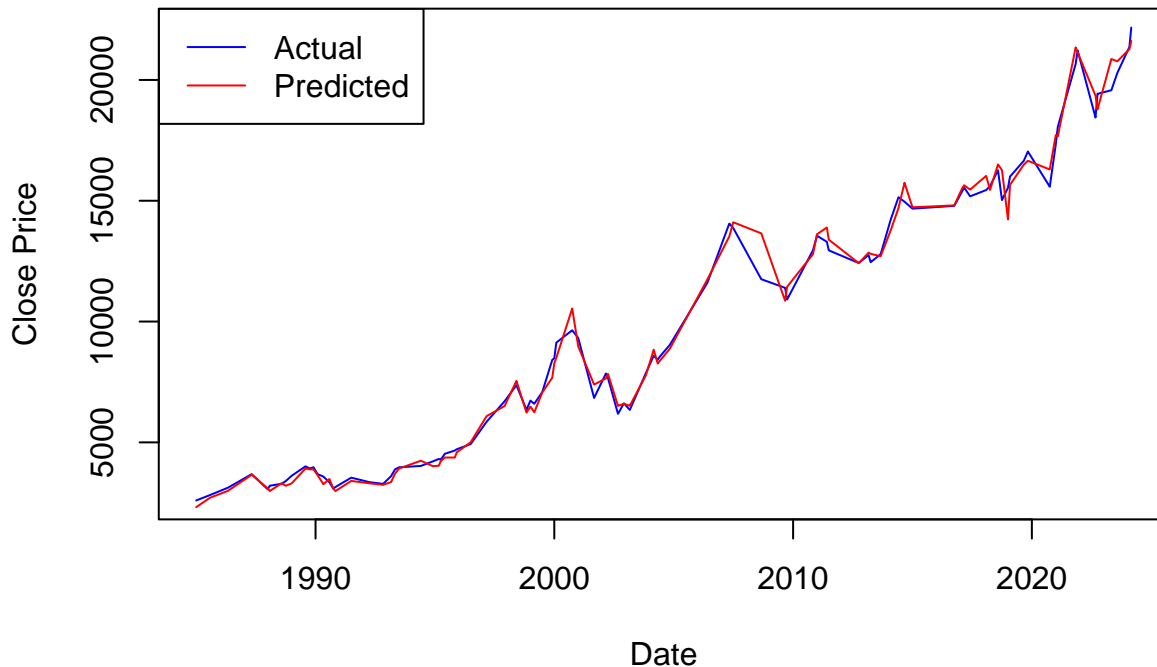
The third plot presents the comparison between actual stock market values and those predicted by the Bayesian model using Markov Chain Monte Carlo (MCMC) simulations. What I'm doing there is separating the dataset into two parts: actual data and predicted data. The advantage of using MCMC is its flexibility to incorporate prior knowledge and quantify uncertainty in predictions through the posterior distributions of parameters in the model. First of all, it's upward-sloping as well. With the fluctuations in different periods, the trend shows a clear increase with time. What's more, the plot shows the convergence of predicted values to actual market data, highlighting the model's effectiveness. One of the strengths of MCMC in this case is its ability to produce a variety of possible future values rather than a single-point estimate. This can reflect the probabilistic nature of financial markets.

However, we have to acknowledge that while the MCMC model has aligned closely with historical data, unexpected market movements, economic changes, and external shocks are difficult to predict and can significantly interfere with the trends.

When I'm constructing the model here, I assume the Ceteris Paribus. This means all the other factors remain the same. The prediction should not be influenced by any other exogenous factors (like the reason for the existence of the error term in the regression in our Model Structure part).

Comparison between the SARIMA model and the MCMC model

SARIMA is a specific type of model used for forecasting time series data that extends ARIMA models to include seasonal effects. This is a very straightforward way to capture the patterns and trends that repeat

4

over fixed periods. However, it always assumes a linear relationship between past values and/or past errors and future values, while sometimes the reality is not.

MCMC is a method for sampling from probability distributions in Bayesian statistical models we covered in the lecture. MCMC is computationally more intensive and can be complex to diagnose and interpret. In general, MCMC methods let us figure out the details of complicated models that are too difficult to solve the posteriors with other regular methods.

In this piece of study, I found out that the SARIMA model and the MCMC model all showed a very similar trend. Overall, the choice between SARIMA and MCMC should be guided by the specific needs of the analysis, the nature of the data, and the critical importance of understanding uncertainty in the forecasted values.

Limitations

Economic Limitations

Predicting the stock market is tough because it can be affected by any news or event, no matter how big or small. Even with advanced statistical methods like Variational Bayesian (VB) and Markov Chain Monte Carlo (MCMC), we're not always spot on the "correct" trends. These Bayesian methods assume based on past patterns, which may not always be the case in rapidly changing economic conditions. "All models are wrong, but some are useful."

Theoretical Assumptions

The models we used assume linear relationships and depend heavily on the assumption that distributions from the models can reflect real-world data. Variational Bayesian (VB) tries to take a shortcut to make things more efficient, but sometimes that can lead us off track. What's more, as I mentioned in the MCMC model analysis part, it digs into the details, but it's like solving a giant puzzle that takes a lot of brainpower and time, and if the pieces don't fit just right, we might not be able to predict the unseen future.

Comparison with the SARIMA Model

After introducing the features of those two different models: SARIMA and Baysians, the comparison process can also create some uncertainties. While Bayesian models have a lot of potential like MCMC, they don't always beat SARIMA when it comes to simply predicting where the stock prices will go. We've got to keep a close eye on both, making sure they're doing their job well as we come across new data and new ways of forecasting.

Appendix

References

Chérief-Abdellatif, B.-E., & Alquier, P. (2018). Consistency of variational Bayes inference for estimation and model selection in mixtures. Electronic Journal of Statistics, 12(2), 2995-3035. https://doi.org/10.1214/18-EJS1475

RStan Development Team. (n.d.). Stanmodel method-vb. rstan. Retrieved April 16, 2024, from https://rdrr.io/cran/rstan/man/stanmodel-method-vb.html

Organisation for Economic Co-operation and Development. (2024). Infra-annual labour statistics [Data set]. Retrieved April 18, 2024, from https://data-viewer.oecd.org?chartId=492d0147-1557-4d23-bb66-79395de80775

Organisation for Economic Co-operation and Development. (2024). Production, sales, work started and orders [Data set]. Retrieved April 18, 2024, from https://data-viewer.oecd.org?chartId=f8c73af7-7711-41aa-a84e-e5ae1f6f2126

Organisation for Economic Co-operation and Development. (2024). Financial market [Data set]. Retrieved April 18, 2024, from https://data-viewer.oecd.org?chartId=36e31dd7-9185-4764-bfd5-b151fa20c29c

Organisation for Economic Co-operation and Development. (2024). Consumer price indices (CPIs, HICPs), COICOP 1999 [Data set]. Retrieved April 28, 2024, from https://data-viewer.oecd.org?chartId=9cef7bd5-43eb-4781-909a-b6e7be4f0e11

Yahoo Finance. (2024). TSX index historical data [Data set]. Yahoo! Retrieved April 18, 2024, from https://ca.finance.yahoo.com/quote/%5EGSPTSE/history

Repo Link:

https://github.com/ngggDuy/STAT-447-Project/tree/EdwardBaiyangti-patch-1

Link to the codes for my project:

https://github.com/ngggDuy/STAT-447-Project/blob/EdwardBaiyangti-patch-1/the%20codes%20for%20my%20project.pdf