The Impact of COVID-19 on the Gross National Income of Canada

Daniel J. Polson

December 10, 2021

Introduction

The GNI (Gross National Income) of a country is a measurement used for tracking a countries wealth, given by the total money earned by all people and businesses in the country. In early 2020, the GNI of Canada took a large hit, which is theorized to have been caused by the global pandemic, COVID-19. The goal of this paper is to use ARIMA models to forecast the GNI of Canada given all the data up until the pandemic, and compare it to the actual data.

Analysis

DATA: Our data gives the year, the quarter, and the value of the GNI (at market value) in the given quarter, in units of \$ 1000000 CAD. The data spans from the first quarter of 1961 to the third quarter of 2021, for a total of 242 data points. Figure 1 displays this initial time series. Our analysis will use all points from the first quarter of 1961 to the fourth quarter of 2019 (i.e, all data points from before COVID-19). Call this time series x_t, so that x_i represents the value of the GNI at quarter i.

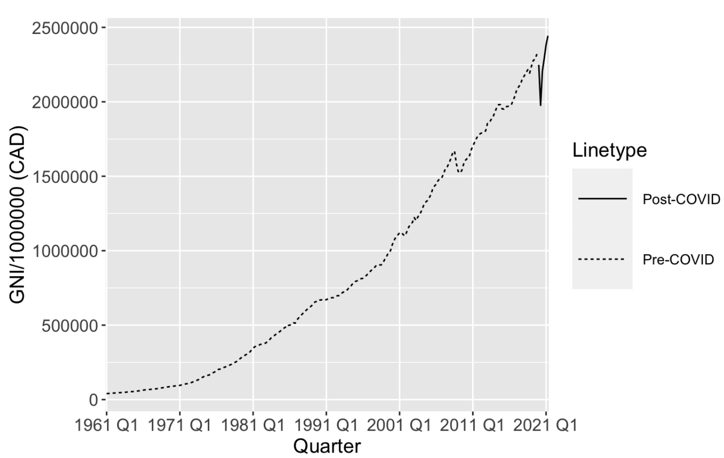

Figure 1: Canadian GNI over time

TRANSFORMATIONS: Since we are using the ARIMA family of models, to begin our analysis, we require our time series to be stationary. We perform a log transformation to stabilize the variance, and a second difference to deal with possible correlations. Then, our transformed time series y_t is given by:

$$y_t = \nabla^2 \left(\log\left(x_t\right)\right) \tag{1}$$

Figure 2 shows the result of this transformation. This transformed time series does not appear to deviate from any of the rules of stationary time series, so henceforth we will assume that y_t is stationary.

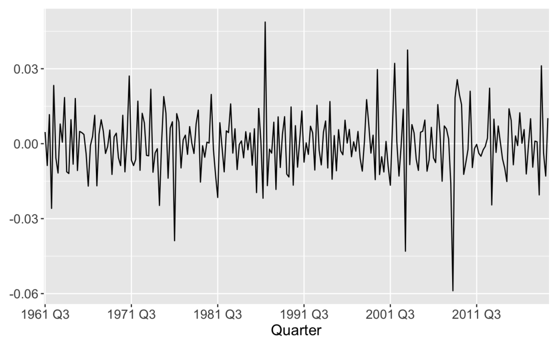

Figure 2: Transformed time series y_t over time

DIAGNOSTICS: To check for any obvious ARMA(p, q) models to fit to y_t, we plot the sample autocorrelation and partial autocorrelation of y_t. Figure 3 shows these plots. It appears there may be a cut off in autocorrelation at lag 1 or lag 2, which would imply that an MA(1) or MA(2) model, respectively, would be the best fit for y_t. The partial autocorrelation does not appear to cut off at any lag, so it is probable an AR(p) model will not be the best fit for y_t. Furthermore, there is no sign of seasonal trends in either plot, so we will not use seasonal ARIMA models. We will test both MA models as well as a number of other ARMA(p, q) to find a model which best fits y_t.

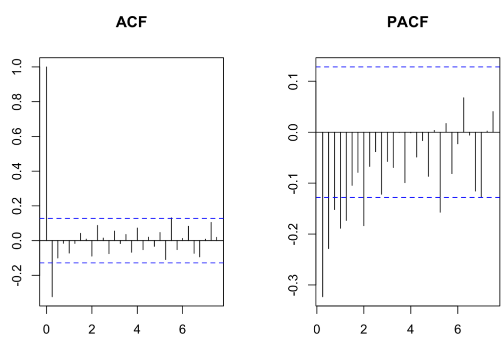

Figure 3: Sample ACF (left) and PACF (right) of y_t for various lag values

MODEL FITTING/SELECTION: In the model fitting process, we tried to fit a number of ARMA(p, q) models, including the ones we hypothesized in the diagnostics section to y_t. Table 1 contains the goodness-of-fit statistics AIC (Akaike

Information Criterion), AICc (Akaike Information Criterion, corrected), and BIC (Bayesian Information Criterion) for the best fitting of these models.

Model	**AIC**	**AICc**	**BIC**
MA(1)	-6.094958	-6.094884	-6.065425
MA(2)	-6.173816	-6.173594	-6.129517
ARMA(1,1)	-6.184798	-6.184576	-6.140499
ARMA(1,2)	-6.176253	-6.175807	-6.117188
ARMA(2,2)	-6.167712	-6.166965	-6.093880

Table 1: Goodness-of-fit criteria of various models for y_t

We want to choose a model which best minimizes these three goodness-of-fit criterion. Clearly, since the ARMA$(1, 1)$ minimizes all three criterion, we choose this model for y_t. The fitted model is then given by:

$$(1 - 0.3486_{(0.0689)}B)y_t = (1 - 0.9362_{(0.0252)}B)w_t \tag{2}$$

However, we want the model written in terms of our original time series x_t. Accounting for the transformations done in the section above, we can write our model as:

$$(1 - 0.3486_{(0.0689)}B)\nabla^2(\log(x_t)) = (1 - 0.9362_{(0.0252)}B)w_t \tag{3}$$

It remains to check that all assumptions of the ARIMA model hold for this fitted model. Figure 4 shows a time-series plot of the standardized residuals for the model, a sample autocorrelation function plot of the residuals, a normal-QQ plot of the standardized residuals, and a plot of the p-values for the Ljung-Box statistic. Aside from a possible outlier just before 2010, the first three plots seem to indicate that the residuals are gaussian white noise, as desired. Furthermore, the last plot shows that there is a high probability that the residuals are uncorrelated for each lag value. Therefore, all of the required assumptions hold, and we choose the model from (3) for the data.

Results

Using our chosen model from the previous section, we will make predictions for the next ten quarters. To do so, we implement the forecast() function in R, which uses finite-history prediction to make forecasts. Figure 5 shows the predicted values from our forecast, as well as 80% and 95% confidence intervals for the predicted values.

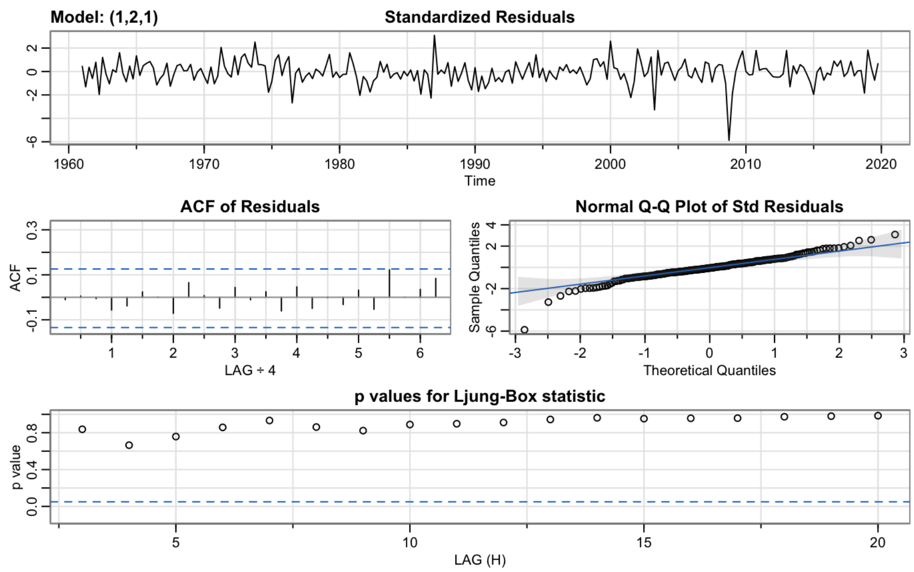

Figure 4: Residual analysis plots for model (3)

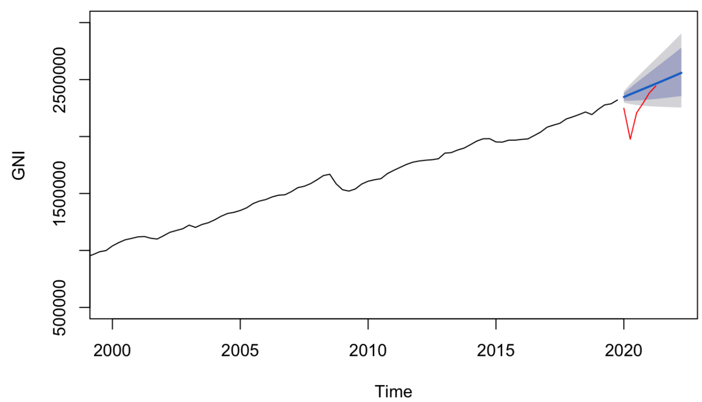

Figure 5: Predicted GNI (blue) versus Actual GNI (red)

Conclusion

Figure 5 presents a number of interesting observations. Firstly, the dip in early 2020 is well outside of the error margin, implying that the dip caused by COVID-19 was significant. Furthermore, the actual data (red line) appears to be converging to the predicted data (blue line) which only used data from before the pandemic. This implies that though COVID-19 caused a large dip in the GNI initially, it may end up having little to no long term effect.

Appendix

See the following pages for R code used.

References

[1] Statistics Canada. Gross domestic income, gross national income, and net national income, doi: 10.25318/3610012201.

[2] David S. Stoffer Robert H. Shumway. *Time Series Analysis and Its Applications*. Springer Science+Business Media, 233 Spring Street, New York, NY 10013, USA, 2011.