PREDICTING GDP GROWTH USING AR MODELS

DANIEL LEWIS

1. Project motivation

The question of predicting future economic growth is a classical problem for econometricians in academia, industry and government. Autoregressive (AR) models are well-suited to this task. For a dataset consisting of Gross Domestic Product (GDP) figures for each of the 50 US states over the last X years, we verify that the AR(4) model gives the best predictions, as theory suggests it should. With this confirmed, we also find it informative to train our models on the periods prior to the 2008 "Great Recession" and the 2020 Covid-19 pandemic, to see how great the disparity is between our forecasted growth and actual growth post-slowdown.

2. The dataset

The data used in this project was obtained from the Federal Reserve Bank of St. Louis [FRE22]. The dataset consists of quarterly data on the seasonally adjusted annual rate of the all industry GDP total, measured in millions of dollars, for each of the 50 US states, together with the District of Columbia. The data stretches from 2005 Q1 to 2022 Q2.

3. Autoregressive models

Recall that if we have only a small number of regressors then the linear model works well for continuous outcomes, and we can use Least Squares. Predicting GDP growth is a prototypical example.

The simplest autoregressive (AR) model is that with only a single lag, the AR(1) model, in which we regress Y_t on $\{1, Y_{t-1}\}$:

$$Y_t = \alpha + \rho_1 Y_{t-1} + \epsilon_t. \tag{AR(1)}$$

We consider also the autoregressive model with 4 lags, i.e. the AR(4) model, where we regress Y_t on $\{1, Y_{t-1}, Y_{t-2}, Y_{t-3}\}$:

$$Y_t = \alpha + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \rho_3 Y_{t-3} + \rho_4 Y_{t-4} + \epsilon_t. \tag{AR(4)}$$

We can then produce predictions iteratively:

$$\begin{aligned}
\hat{Y}_{T+1} &= \alpha_{LS} + \hat{\rho}_1 Y_T + \hat{\rho}_2 Y_{T-1} + \hat{\rho}_3 Y_{T-2} + \hat{\rho}_4 Y_{T-3}, \\
\hat{Y}_{T+1} &= \alpha_{LS} + \hat{\rho}_1 Y_T + \hat{\rho}_2 Y_{T-1} + \hat{\rho}_3 Y_{T-2} + \hat{\rho}_4 Y_{T-3}, \\
&\vdots
\end{aligned}$$

Here α_{LS} and $\hat{\rho}_i$ denote the Ordinary Least Squares (OLS) predictions of the coefficients α and ρ_i.

Date: November 14, 2022.

Theory suggests the AR(4) model as the best predictor of economic growth. A priori, an AR(4) model should outperform a model with fewer lags, since it approximates the data generating process (DGP) better. Moreover, an AR(4) model should outperform a model with more lags since the coefficients will be estimated with more precision.

The principal goal of this project is to verify that the AR(4) model does indeed outperform other specifications when applied to our real world dataset (described in Section 2).

4. Performance metrics

We use *Mean Squared Forecasting Error (MSFE)* as our key metric in comparing models' predictions. Recall that the MSFE of a prediction $\tilde{Y}_t$ of a random variable Y_t is defined as the expectation of the squared difference:

$$\mathrm{MSFE}(Y_t, \tilde{Y}_t) := E[(Y_t - \tilde{Y}_t)^2].$$

Thus a lower MSFE indicates a better prediction.

5. Exploratory data analysis

Using R [RFfSC22] in RStudio [RT22], we plot the trend in GDP for a given state using our dataset. The patterns of GDP growth in the states of Arizona and California, included in Appendix 10, are indicative of the general trend. Namely, GDP grew year on year, with the exception of the periods immediately following the onset of the "Great Recession" of 2008 and the Covid-19 pandemic of 2020.

Once we have computed predictions using the AR models, we can append these to the existing plots, including error bars to quantify the uncertainty in each prediction.

6. Plan of analysis

Following the example of [HAGS21, §14.3], we subset the data to focus on the state in question – for us, the focus will naturally be on Arizona – and apply the R command `ar.ols` from the package `stats` to estimate the model. An important point is to input growth rate rather than raw GDP into the AR model.

7. Model development

8. Results

9. Conclusion

References

[FRE22] U.S. Bureau of Economic Analysis FRED Federal Reserve Bank of St. Louis, *Gross Domestic Product: All Industry Total* (2022), `https://fred.stlouisfed.org/series/ALNQGSP`. ↑2

[HAGS21] Christopher Hanck, Martin Arnold, Alexander Gerber, and Martin Schmelzer, *Introduction to Econometrics with R*, 2021. ↑6

[RFfSC22] R Core Team R Foundation for Statistical Computing (Vienna), *R: A language and environment for statistical computing* (2022), `https://www.R-project.org/.`. ↑5

[RT22] RStudio Team, *RStudio: Integrated Development Environment for R* (2022), `http://www.rstudio.com/`. ↑5

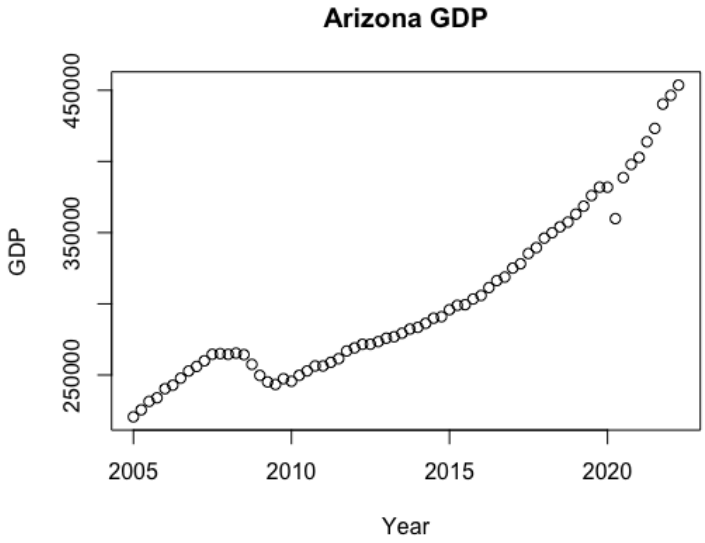

FIGURE 1. Arizona GDP

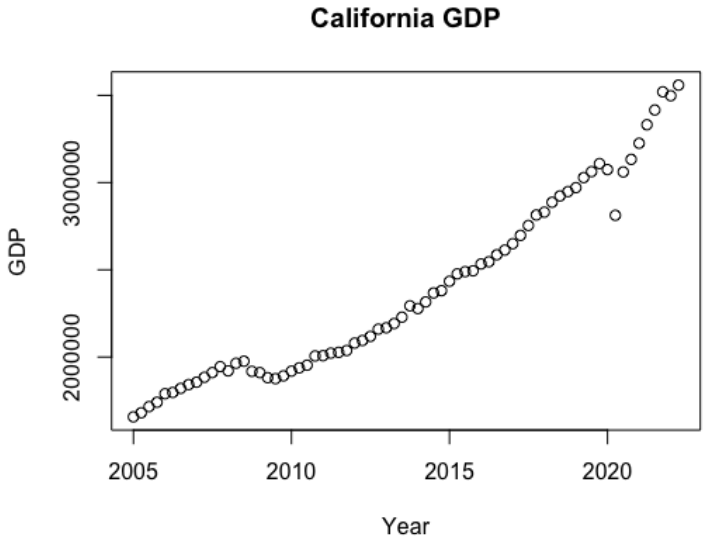

FIGURE 2. California GDP

10. APPENDIX

We include here the results of our exploratory data analysis for the states of Arizona and California. Note the increasing trends, bar exceptions (see Section 5).