PREDICTING GDP GROWTH USING AR MODELS

DANIEL LEWIS

Date: December 7, 2022.

1. Project motivation

The question of predicting future economic growth is a classical problem for econometricians in academia (see e.g. [HY22]), industry ([Bac22]) and government ([OfNS22]). 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 whether the AR(4) model gives the best predictions, as theory suggests it should. 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. Outline

In Section 3 we describe our dataset and where it was found.

Section 4 outlines the autoregressive models we are to compare.

Section 5 describes the performance metrics we use in our analysis: naïve forecasting error, mean squared forecasting error and R^2.

We then present our explanatory data analysis in Section 6, and our plan of analysis in Section 7.

Finally, we provide our results in Section 8, and highlight our conclusions in Section 9.

We also include three appendices: the first, Appendix I, describes the reproducibility of our results, while the second, Appendix II, presents the graphical results of our exploratory data analysis. Appendix III considers our "side quest" of imagining economic growth had the 2020 Covid-19 pandemic not occurred.

2.1. **Acknowledgments.** Many thanks to Ami Ichikawa for helpful discussions.

3. 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: 70 quarters in all.

4. 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:

$$\hat{Y}_{T+1} = \alpha_{LS} + \hat{\rho}_1 Y_T + \hat{\rho}_2 Y_{T-1} + \hat{\rho}_3 Y_{T-2} + \hat{\rho}_4 Y_{T-3}, \tag{4.1}$$

$$\hat{Y}_{T+2} = \alpha_{LS} + \hat{\rho}_1 \hat{Y}_{T+1} + \hat{\rho}_2 Y_T + \hat{\rho}_3 Y_{T-1} + \hat{\rho}_4 Y_{T-2}, \tag{4.2}$$

$$\vdots$$

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

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 3).

4.1. **Modeling assumptions.** As described above, autoregressive models are reliant upon the ordinary least squares (OLS) regression of Y_t on lagged variables Y_{t-i}. Thus, for our models to perform effectively, we require the standard OLS assumptions to hold:

1. The error term ϵ_t is uncorrelated with all regressors.
2. Errors are uncorrelated: for any $t_1 \neq t_2$, we have $E\epsilon_{t_1}\epsilon_{t_2} = 0$.
3. The "full rank" assumption: no regressor is a linear function of other regressors.
4. Fourth moments exist: $E[Y_t^4] < \infty$ for all t.

Assumption 1 holds since we only ever regress Y_t on past data Y_{t-i}. As Y_t contains ϵ_t, it would not be valid to assume $E[\epsilon_t \mid Y_1, \ldots, Y_t] = 0$.

Assumption 2 is a fair assumption for time series data: there is no reason the error in regressing, say, 2021Q3 on 2021Q2 data should be correlated with the error in regressing 2021Q2 on 2021Q1. In fact, errors are independent and identically distributed.

Since we work with GDP figures, assumption 4 poses no difficulty: production is finite, so the expectation of a regressors fourth power will be finite as well.

Assumption 3 is clearly violated: equation (4.1) demonstrates how each regressor is a linear function of prior regressors. Nonetheless, autoregressive models provide consistent estimators under mild regularity conditions [Han22, Theorem 14.29].

5. 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.

We also include in our presentation a naïve "forecasting error" (FE), computed straightforwardly as

$$\mathrm{FE} = Y_t - \tilde{Y}_t.$$

This has the advantage of being quick and easy to compute, but is not as refined a metric as MSFE. A FE lesser in absolute value will indicate a better prediction.

Finally, we include the R^2 from our regressions, defined as

$$R^2 = \frac{n^{-1}\sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2}{n^{-1}\sum_{i=1}^{n}(y_i - \overline{y})^2},$$

where $\hat{y}_i = x_i'\hat{\beta}$ and $\overline{\hat{y}} = n^{-1}\sum_{i=1}^{n}\hat{y}_i$. The R^2 values computed will indicate how much of the trend in GDP is explained by the autoregressive model. A higher R^2 is preferable, but can sometimes be misleading. Low R^2 may signify an issue.

6. 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 II, 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.

7. 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: failing to do so, we notice unrealistic results, such as a projected fall in GDP when given an increasing trend.

The first step in our analysis is to train each model on the first 69 data points, and compare (via naiïve FE) the predicted GDP figures for 2022 Q2 against the true figures provided in our dataset. Then, using the real data rather than this predicted 2022 Q2 figure, we predict ahead to 2022 Q3, Q4 and 2023 Q1. No comparison is possible for this out-of-sample prediction.

Our second step is to train the model on the first 50 datapoints in our sample, and use this to make predictions for the remaining 20 datapoints. With these predictions in hand, we may compute an estimate of the MSFE for these predictions for each model, as well as providing the R^2.

Finally, we address our "side quests", comparing predicted growth in 2008 and 2020 against the actual figures marking these downturns.

8. Results

In the first step outlined in Section 7, we obtain the following results: see Table 8.1.

One should note in particular the strong performance of the AR(4) model. However, the AR(3) model appears to have performed better in this instance, producing an FE smaller in absolute value.

However, this is very weak experimental evidence as to which model performs best: we have so far produced only a single prediction, so one should not be swayed to any conclusion.

-	Forecast	FE	Runtime (s)
True GDP	453602.3	-	-
AR(1)	451282.5	-2319.802	0.93
AR(2)	451375.5	-2226.806	1.05
AR(3)	453340	-262.2576	1.07
AR(4)	453943.2	340.8569	0.99
AR(5)	454395.6	793.3188	1.05
AR(6)	454404.8	802.4994	0.92
AR(8)	452720.7	-881.6492	1.11
AR(12)	458074.9	4472.624	1.24
AR(16)	460015.5	6413.18	0.96

TABLE 8.1. Step one: we compare the predicted Arizona GDP figures for 2022 Q2 against the true figures provided in our dataset.

-	MSFE	R^2	Runtime (s)
AR(1)	2979313641	0.4764882	0.47
AR(2)	1042557581	0.3654467	0.50
AR(3)	1334502169	0.3573387	1.00
AR(4)	1267756968	0.3614263	0.79
AR(5)	2090562081	0.3576215	0.91
AR(6)	1772792756	0.364209	1.00
AR(8)	2619002637	0.4082417	1.04
AR(12)	2952219077	0.485657	2.21
AR(16)	395146580	0.5845539	1.48

TABLE 8.2. Step two: we train each model on the first 50 data points in our sample and predict the remaining 20 data points.

The next substep in our analysis is to predict ahead to 2022 Q3, Q4 and 2023 Q1. Given we have no real data for comparison (no such figures were available at the time we accessed the dataset), there is no value to be gained in presenting the raw predictions. However, we find it informative to add these predictions to the visualizations included in Appendix II, colored red to distinguish these predictions from real data.

For step two as outlined in Section 7, we observe the following (see Table 8.2): while the AR(4) model appears to perform somewhat well, it is outperformed both by the AR(2) model on one hand, and the AR(16) model on the other. This runs contrary to theory: recall that an AR(4) model should outperform a model with fewer lags since it approximates the data generating process better, and should outperform a model with more lags since the coefficients will be estimated with more precision.

Moreover, this unexpected result is not a one-off: we observe the same trend when applying our models to the GDP data for the state of California. Moreover, the higher R^2 values for the AR(16) model suggests it fits the data better and so is adept at predicting future growth.

A possible explanation for this surprising result is that it does not suffice to just train the model on 50 data points to produce 20 predictions. A more refined approach, such as k-fold cross-validation, may point to the AR(4) model as the best predictor. However, such an analysis falls beyond the scope of this paper.

Finally, we comment briefly on our side quests. The first (training on data prior to the 2008 recession) proved inconclusive: perhaps training our model on just eleven data points is insufficient.

The second additional task, to predict economic growth had the Covid-19 pandemic not occurred, proved to be more interesting: see Appendix III. Note in particular that our AR(4) model predicts continued economic growth, without the 2020 slump. However, by the year 2021 actual growth has caught up with these predictions, and in 2022 significantly exceeds our model's predictions.

9. Conclusion

Our principal task in this investigation has been to apply autoregressive models to real-world data to predict future economic growth. In particular, we have compared AR(p) models for various numbers of regressors, p. Theory suggests the AR(4) model should provide the best predictions.

However, our experimental evidence points to the AR(16) model as the best predictor. This is indicated by its having the lowest estimated mean squared forecasting error, coupled with a high R^2. This runs contrary to our expectations based on theory: an AR model with more than 4 lags should perform worse than the AR(4) model, since the coefficients will be estimated with less precision.

We explain this surprising result by pointing to the small sample size upon which we make predictions. Averaging the squared differences in 20 predictions provides a reasonable first estimate of MSFE, but a more reliable estimate could be obtained by k-fold cross-validation, which unfortunately falls beyond the scope of this project. Thus, we have insufficient evidence to reject the hypothesis that AR(4) models provide the best prediction of future economic growth. The task warrants further investigation.

References

[Bac22] Daniel Bachman (2022), `https://www2.deloitte.com/us/en/insights/economy/us-economic-forecast/united-states-outlook-analysis.html`. Accessed December 1, 2022. ↑1

[DGC+19] Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, and Noah A. Smith (2019), `https://arxiv.org/pdf/1909.03004.pdf`. Accessed December 4, 2022. ↑10

[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`. ↑3

[HAGS21] Christopher Hanck, Martin Arnold, Alexander Gerber, and Martin Schmelzer, *Introduction to Econometrics with R*, 2021. ↑7

[Han22] Bruce E. Hansen, *Econometrics*, 2022. ↑4.1

[HY22] Yingyao Hu and Jiaxiong Yao, *Illuminating economic growth*, Journal of Econometrics **228 (2)** (2022), 359–378. ↑1

[OfNS22] Office for National Statistics (2022), `https://www.ons.gov.uk/economy/grossdomesticproductgdp/bulletins/gdpmonthlyestimateuk/september2022`. Accessed December 1, 2022. ↑1

[RFfSC22] R Core Team R Foundation for Statistical Computing (Vienna), *R: A language and environment for statistical computing* (2022), `https://www.R-project.org/`.. ↑6

[RT22] RStudio Team, *RStudio: Integrated Development Environment for R* (2022), `http://www.rstudio.com/`. ↑6

10. Appendix I

A few remarks are in order concerning the reproducibility of our results, inspired by [DGC+19].

- Description of computing infrastructure
  All computation was run on the author's Macbook Air with a 1.1 GHz Dual-Core Intel Core i3 processor and 8GB memory.
- Average runtime for each approach
  See Table 8.1 and Table 8.2 in Section 8.
- Details of train/validation/test splits
  See Section 7.
- Corresponding validation performance for each reported test result
  There are not separate validation and test sets in this paper.
- A link to implemented code
  See `https://github.com/danlewis92/GDPproject`.

11. Appendix II

We include here the results of our exploratory data analysis for the states of Arizona and California, updated to include the next 3 predictions of the AR(4) model, plotted in red. Note the increasing trends, bar exceptions (see Section 6).

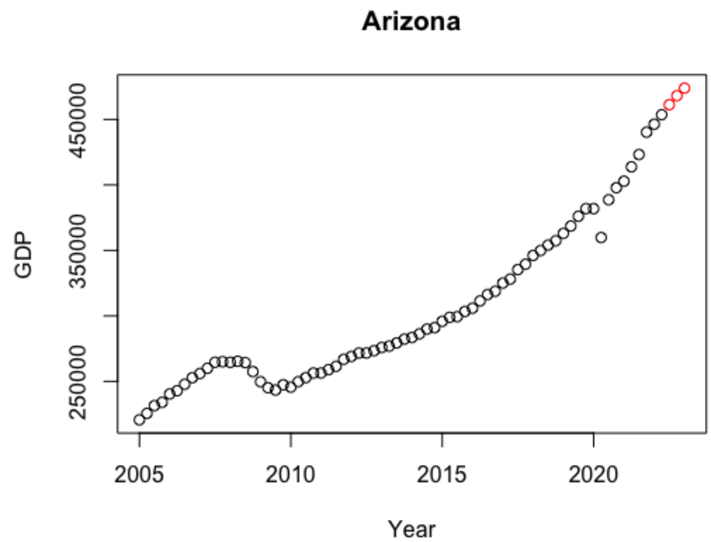

Figure 1. Arizona GDP

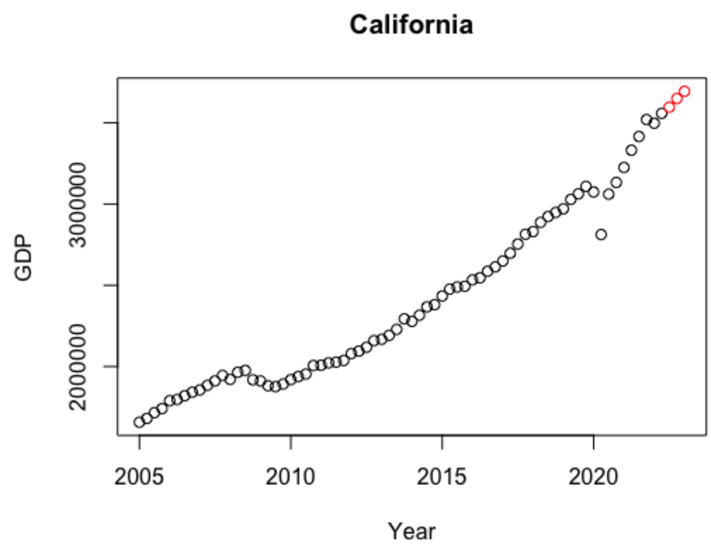

Figure 2. California GDP

12. Appendix III

We include here the results of our side quest predicting economic growth had the Covid-19 pandemic not occurred. Predicted growth is plotted in green.

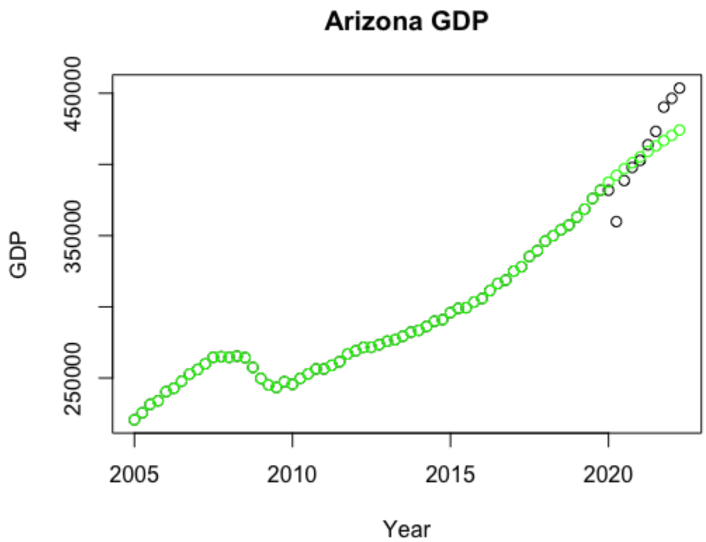

FIGURE 3. Arizona GDP with/without Covid