Statistics 221 – Final Report
Peter Racioppo

Analysis of Global GDP per Capita, 1900 – 2019

Introduction:

At the beginning of the 21st century, the structure of world capitalism is undergoing changes as sweeping and fundamental as those of the Industrial Revolution at the end of the 18th century and the rise of monopoly capitalism during the end of the 19th century. Revolutionary advances in information technology have facilitated a profound internationalization of economic production. Transnational corporations coordinate their activities on a world scale and produce for a world market, while international finance capital scours the globe for new markets and investments. As this process has drawn in ever larger sections of the world over the last four decades, the percent of the world's population involved in industrial production has increased dramatically. Only in the late 1970s or early 1980s did the majority of humanity cease to be agrarian [1]. Such an epochal shift is deserving of careful study. In this paper, I investigate average world income since 1800 in terms of inflation-adjusted gross domestic product (GDP) per capita, as a measure of economic development.

A nation's gross domestic product is a measure of the wealth produced by the working population residing in that country in a year. GDP per capita is a nation's GDP divided by its midyear population. Economic growth and population growth have both been exponential over the last century, but the former far outpaces the latter, so that world GDP *per capita* has also grown exponentially. This report uses world GDP per capita and population data from 194 countries for the years 1800-2019, with projections through 2040, adjusted for inflation and given in terms of 2019 US dollars. Data was taken from gapminder.org [2] which took GDP data from World Bank national accounts [3] and population data from OECD National Accounts data files [4]. GDP and population are among the most carefully-recorded and reliable indicators of national development. To obtain a univariate time series of global GDP per capita, I multiplied each country's national GDP per capita in each year by its population and then summed over all countries to obtain the total global GDP for that year. Dividing by the world population for that year gave the global GDP per capita. World GDP per capita from 1800 to 2019 is displayed in Fig. 1. Growth between about 1800 and 1920 appears quite smooth, but as the left side of Fig. 3 shows, growth rates trend upward and become more volatile during this period. The epoch of monopoly capitalism, beginning around 1900 and characterized by the concentration of production and wealth, the economic domination of monopolies, and the export of capital to the undeveloped countries, corresponds to a period of higher volatility and large drops in GDP per capita, particularly during the first half of the 20th century. One can observe large drops corresponding to World War I and the Spanish Flu, the Great Depression, and World War II. Also visible are the 1980s recession and the Great Recession of 2008. In the following analysis, I focus on this more volatile period from 1900 to 2019. As shown in Fig. 2, the autocorrelation (ACF) of the raw data appears to tail off slowly (left-hand side of the figure), but the partial autocorrelation (PACF) cuts off dramatically after a lag of one year (right-hand side). Thus, a reasonable first guess at a dynamic model is an autoregressive model of lag one (denoted AR(1)).

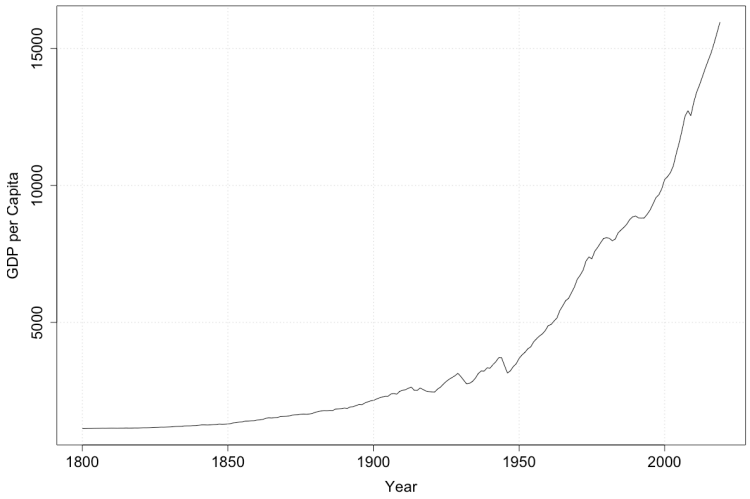

Figure 1. World GDP per capita, 1800-2019 (left), and percent growth rates (right).

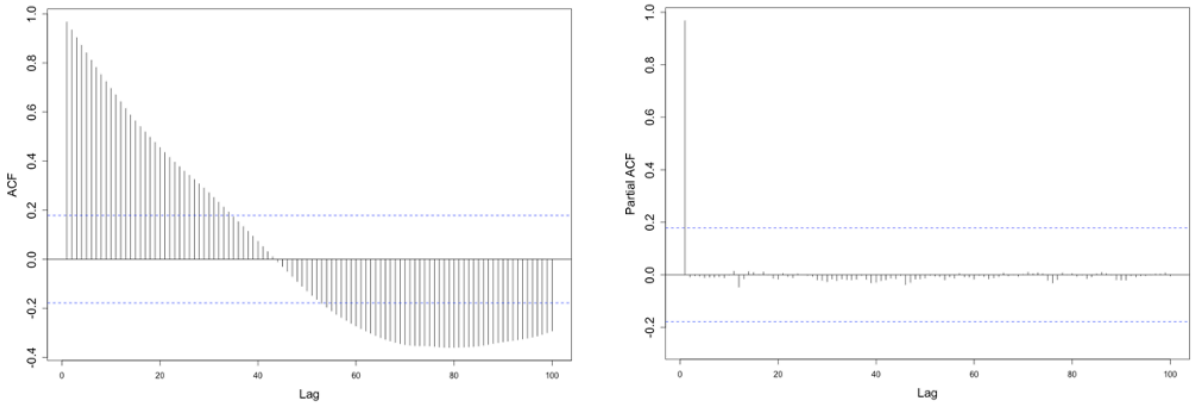

Figure 2. ACF (left) and PACF (right) of raw world GDP per capita data.

When the data can be reasonably assumed to obey dynamics of near-constant growth, and thus to be well-modeled as exponential, the analysis may be simplified by taking a logarithm of the time-series. However, doing so imposes a relationship on the data where none may exist, so we may instead choose to focus on the differenced data, as in the right side of Fig. 1. I will attempt both approaches. Figure 3 (left) displays percent growth rates per year, with a linear trend overlaid. The right side of Fig. 3 displays this data with the linear trend removed.

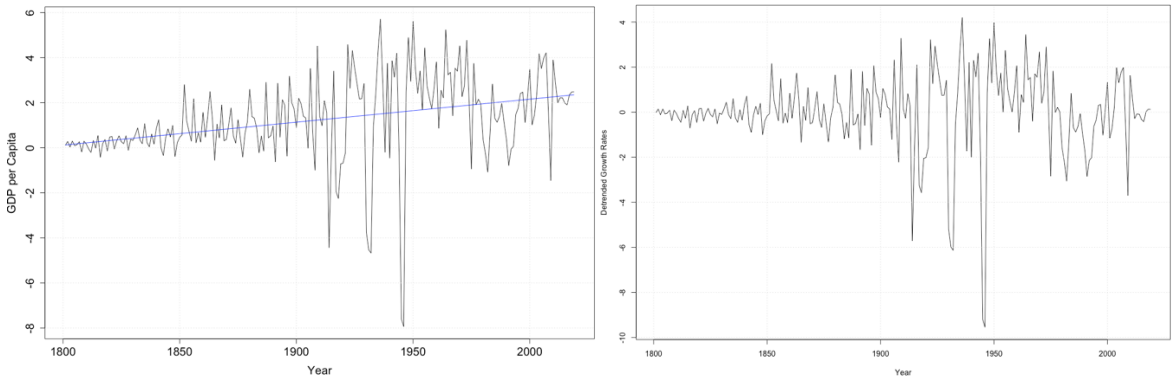

Figure 3. Percent-difference GDP per capita (1800-2019) (left) and detrended (right).

The autocorrelation (ACF) and partial autocorrelation (PACF) of the detrended data are pictured in Fig. 4. Both appear to tail off gradually, which indicates that more differencing may be necessary. Figure 5 displays the periodogram (left) and smoothed periodogram, with a lag of four years (right). The largest peak in the smoothed periodogram occurs at a frequency of 0.7 1/yr, which corresponds to a period of 7.1 years. A second peak occurs at a period of 14.3 years, with a third peak at 56.3 years. Only the 7-year peak lies above its lower confidence bound.

Using the *forecast* library in R to fit a minimum Akaike Information Criterion (AIC) model to the detrended data, gives an ARIMA(0,0,1), or a simple moving average model with lag of one year (MA(1)), displayed alongside the original data in Fig. 6. The residuals (right side of Fig. 6) are similar to the original data, with a marked increase in volatility over time. As shown in Fig. 7 (left), the MA(1) model forecasts a flat line for the period between 2020 and 2039. The AIC is an application of the principle of maximum entropy. Roughly speaking, an uninformative flat line forecast indicates that the ARIMA models are not able to identify generalizable information from the detrended time series. Removing the 7-year seasonal average and fitting an ARIMA model again yields an MA(1) (right side of Fig. 7). In this case, removing the seasonal data actually obscures the trend of increasing volatility, so it may be advisable to scale the subtracted seasonal component by a linear fit to the size of the deviations from the mean.

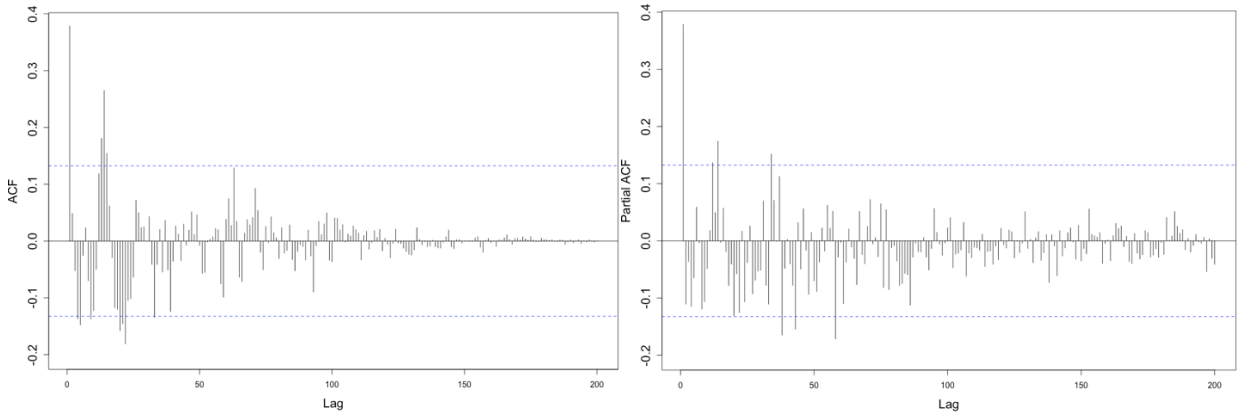

Figure 4. ACF (left) and PACF (right) of the detrended data.

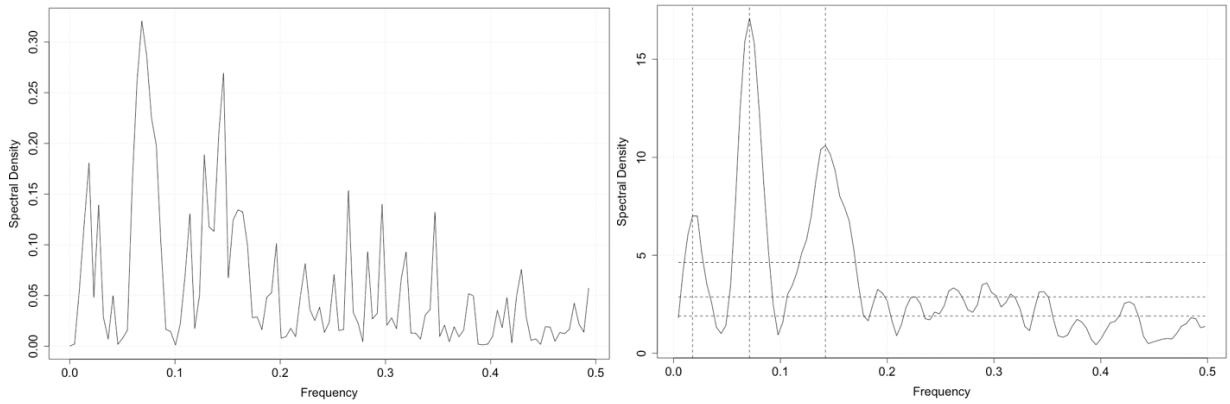

Figure 5. Periodogram of the detrended data (left) and smoothed periodogram with lower confidence bounds (right). Horizontal lines are the lower bounds for the three tallest peaks.

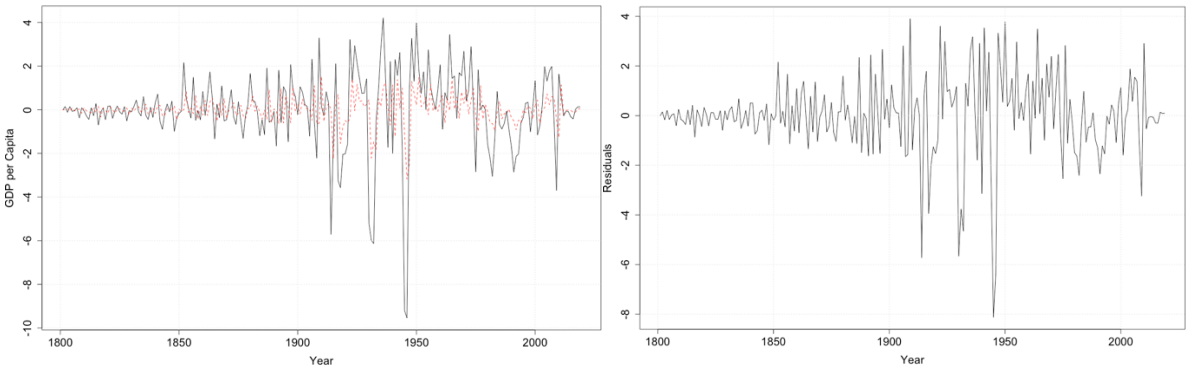

Figure 6. MA(1) fit (left) and residuals (right).

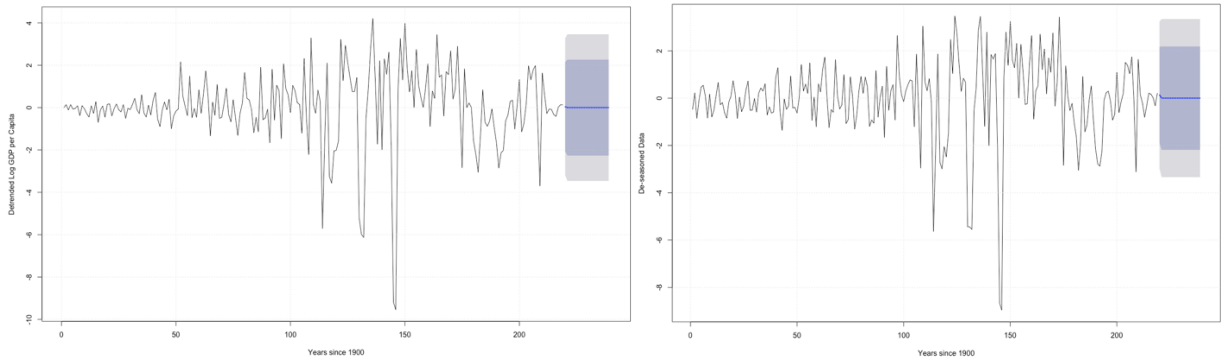

Figure 7. MA(1) forecast of detrended data: 7-year period not removed (left) & removed (right).

Taking a logarithm of the data in Fig. 1 rather than immediately differencing produces the yearly (exponential) growth rate, displayed on the left-hand side of Figure 8. One can observe that the growth rate data still appears somewhat exponential, especially between 1800 and 1900. Fitting a linear trend to the data between 1900 and 2019 gives a GDP per capita doubling time of about 40 years, and subtracting this linear trend produces yearly detrended log-GDP per capita data, displayed on the right side of Fig. 8, which can be interpreted as the yearly deviation from the normal growth rate. The three dramatic dips between 1920 and 1945, corresponding to World War I / Spanish Flu, the Great Depression, and World War II, are now much more visible, and can be seen to come in roughly equal 15-year intervals and to be of similar size and duration. Examining the periodogram of the detrended data, displayed in Fig. 9, we can identify two peaks, corresponding to periods of about 60 years and 15 years. Confidence intervals were, for the 1st peak: [0.05, 7.18], and for the 2nd peak: [0.005, 0.66]. Both peaks are thus significant, and no part of the spectrum with frequencies greater than 0.1 (that is, periods lower than 10 years), has values in this range. However, both peaks in the periodogram appear quite wide, and smoothing the periodogram with a lag of four years, as in the right side of Fig. 9, almost completely removes the existence of the 15-year period. It seems that all that can be confidently said is that almost all of the energy in the spectrum is concentrated at periods above 10 years.

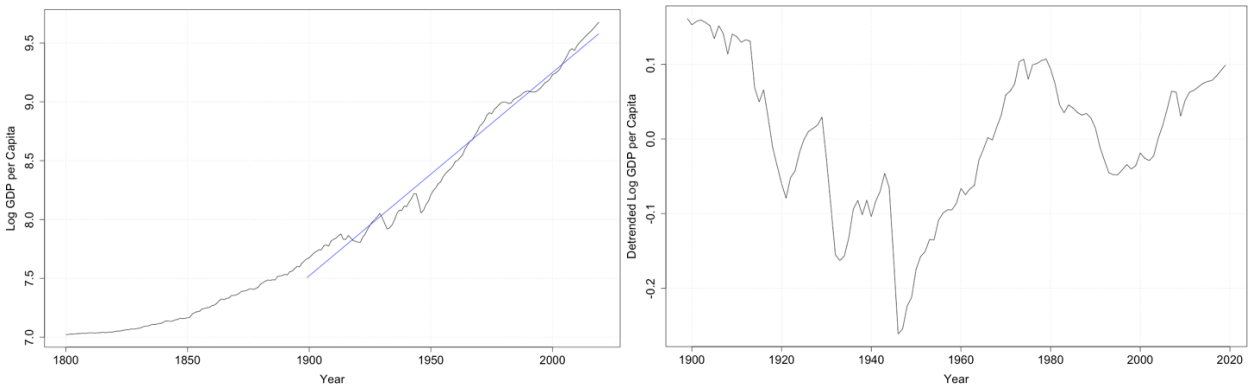

Figure 8. Logarithm of GDP per capita data (left) and detrended log data (right).

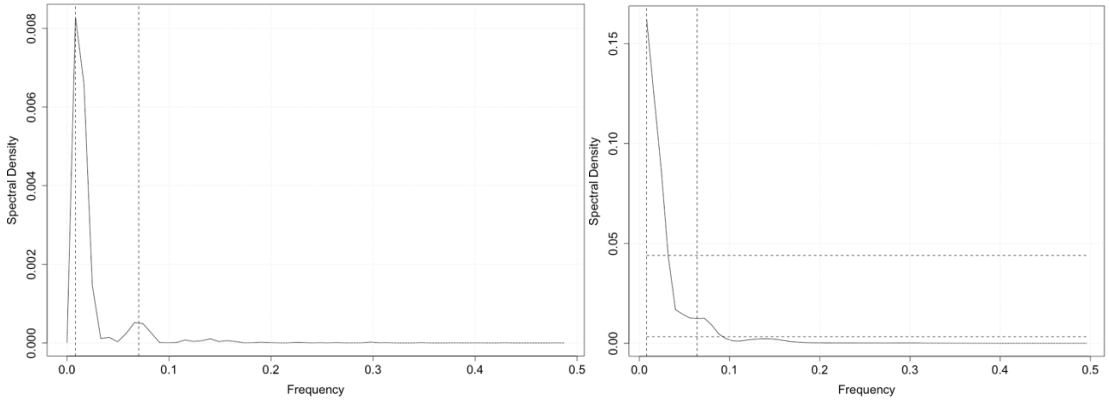

Figure 9. Periodogram (left) and smoothed periodogram with lower confidence bounds (right).

Computing autocorrelations and partial autocorrelations of the detrended data (Fig. 10), we again find that the ACF (left side of Fig. 10) trails off gradually, which indicates that we should consider differencing the time series, while the PACF (right side) again cuts off, but after a lag of two years rather than one. Not surprisingly, the AIC of pure AR fits (left side of Fig. 11) is minimized at a lag of two years. Fitting ARMA(2,0) and ARMA(2,1) models, we find that the sum of square residuals is lower for the latter model. Forecasting 10 years into the future (to 2029), the ARMA(2,1) predicts a slight downward trend, with large margins of error, as pictured in the right hand side of Fig. 11.

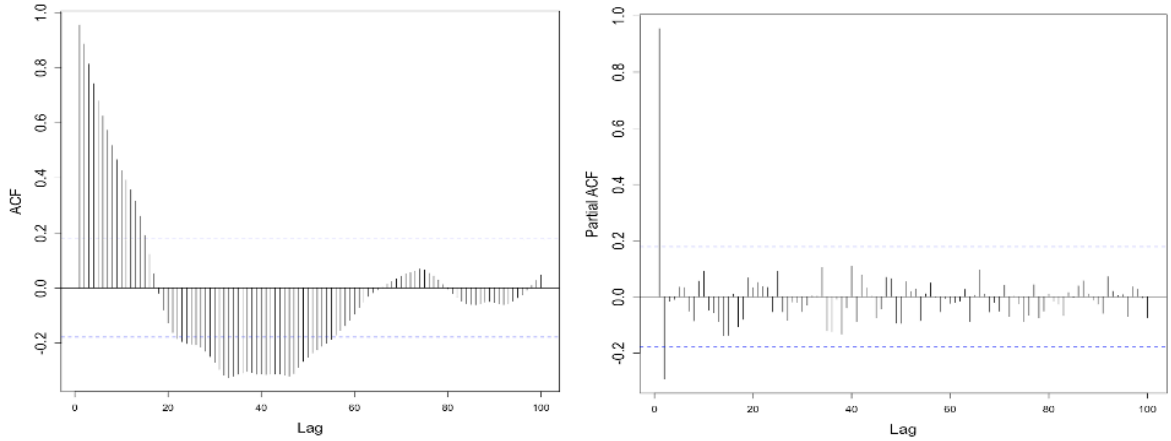

Figure 10. ACF (left) and PACF (right) of the detrended data.

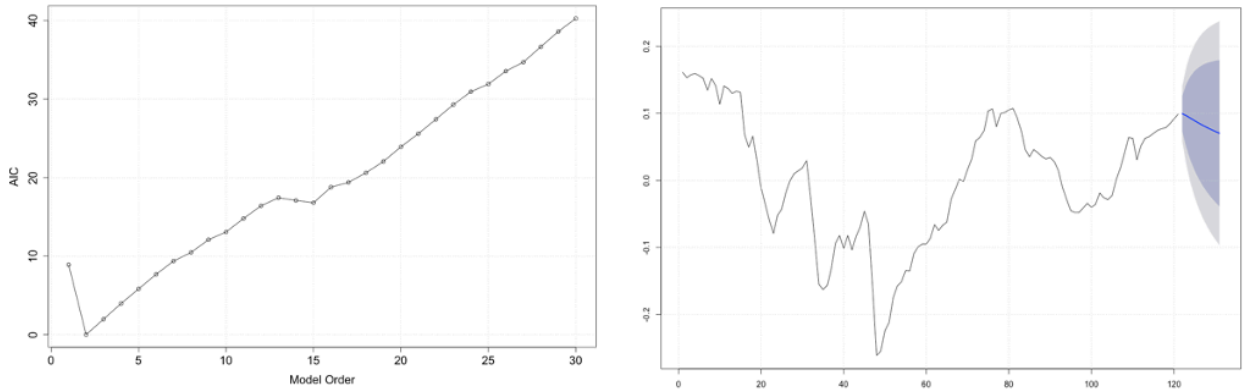

Figure 11. AIC of AR fits to the detrended data (left) and 10 year ARMA(2,1) forecast (right).

On the other hand, using the *forecast* library in R to fit a minimum AIC model to the detrended data, gives an ARIMA(0,1,1), whose predictions are displayed alongside the data in Fig. 12 (left side). The residuals (right side of Fig. 12) appear very noisy, with no obvious trends. As shown in Fig. 13 (left), the ARIMA(0,1,1) forecasts a flat line for the period between 2020 and 2039. Note that the selection of the ARIMA(0,1,1) model agrees with the selection of the MA(1) fit on the differenced data.

I attempted to improve the fit by removing the possible period at 60 years (removing the 15-year period had almost no effect). Fitting again on the de-seasoned data, the best model (using the AIC criterion) was an ARIMA(2,1,2). The forecasts using this model (shown in Fig. 13, right) were very similar to those of the ARIMA(0,1,1): an almost completely flat line, with confidence intervals of approximately the same size. It's interesting to note that the possible 60-year cycle we identified roughly corresponds to the period of a so-called "Kondratiev wave," an hypothesized cycle in world-economy, and the 15 year period is also similar to the length of a "business period" of which there is some speculation in the literature. However, the importance or existence of these waves is not well supported by our analysis, and their inclusion in the ARIMA modeling had little effect on our forecasts. Indeed, speculations on economic waves are not accepted by the majority of academic economists.

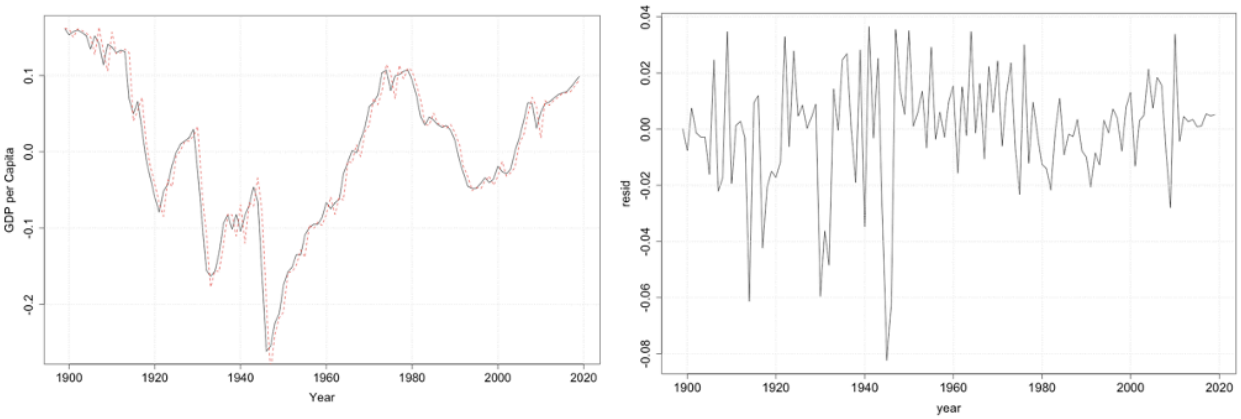

Figure 12. ARIMA(0,1,1) predictions (left) and residuals (right).

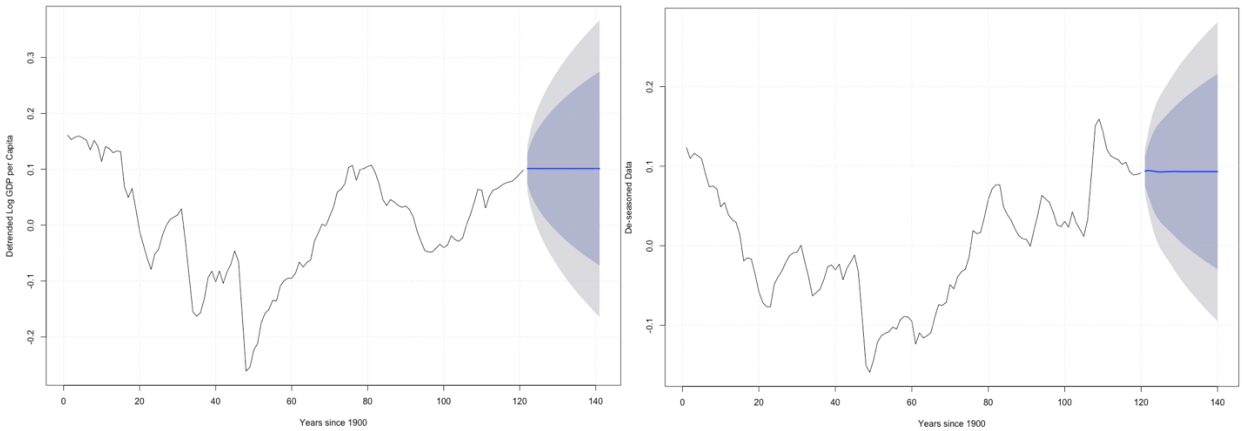

Figure 13. Twenty year detrended ARIMA(0,1,1) forecast (left) and ARIMA(2,1,2) (right).

Figure 14 (left) displays forecasts using the ARIMA(0,1,1) model, transformed back into the original time domain. As shown in the figure, the projections are only slightly less optimistic than the official World Bank projections. The 95% confidence intervals (also shown in Fig. 14) predict that an economic downturn will last no more than about 5 years before a resumption of exponential growth. On the other hand, fitting an ARIMA model to the raw data (Fig. 14, right), we find the best fit, an ARIMA(0,2,2) gives similar projections, but with narrower confidence bands. The higher-order differencing in this model can be seen as an alternative to the use of the logarithm transformation.

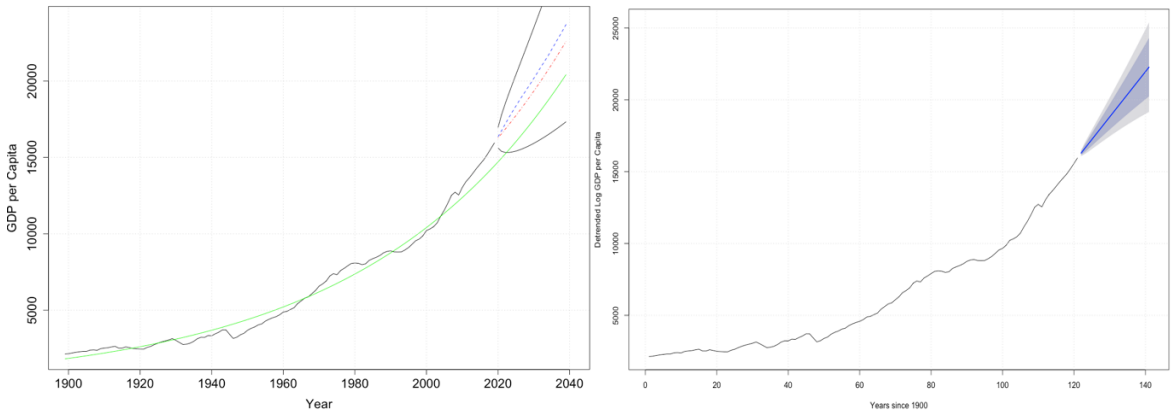

Figure 14. Left Hand Side: Forecasts in the original time domain. In the figure, the upper (blue) dotted line uses the official forecast data, the lower (red) dotted line is the ARIMA(0,1,1) forecast, and the solid lines denote the 95% confidence interval. The green solid line is an exponential fit of the data since 1900. Right Hand Side: An ARIMA(0,2,2) fit of the raw data.

Conclusion:

Evidently, a more complex model is necessary if we want to reproduce the nonlinear and oscillatory dynamics at play in world economy. Recessions are clearly somewhat periodic, but their frequency and severity appear to be chaotic. ARMAX models may be a useful means of capturing the volatile character of economic fluctuations.

Perhaps most critically, we should extend the analysis to include more variables. By averaging per capita GDP over the entire world's population, we lose valuable information on the distribution of wealth. In fact, even the World Bank / UN data does not account for inequality *within* countries (a major factor!). Figure 15 displays income variance across the 194 countries for each year, using the same data set. Note that this data does not account for inequality *within* countries and therefore does not fully capture global inequality. However, there are clear correlations with the mean income data in Fig. 1, including most notably, the three dips between 1920 and 1945.

State-space methods may be a helpful means of considering the effect of wealth distribution. In the post-World War II period, for example, rapid growth in the developed countries caused the distribution of world income versus population to separate into a roughly bimodal distribution. The globalization of production has since reversed this process, with the two peaks of the histogram remerging into a unimodal distribution. Adding one extra state-variable would allow us to account for the existence of a bimodal wealth distribution (e.g. by separately tracking the peak of each mode), which may account for some of the observed oscillatory dynamics.

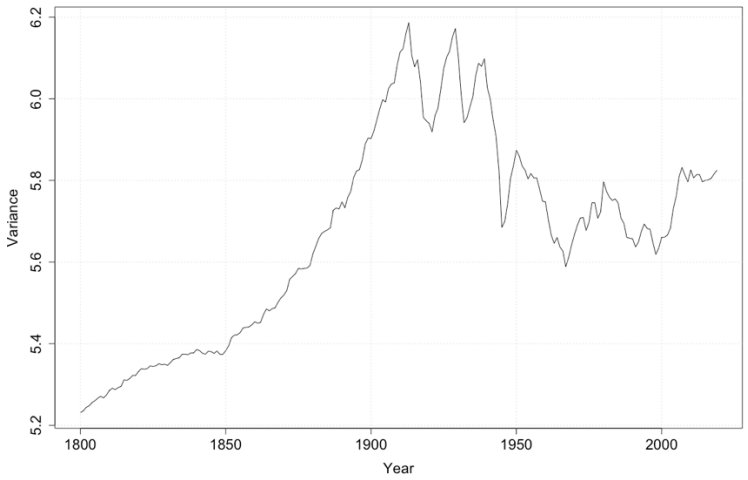

Figure 15. Income variance across countries, 1800-2019.

Citations

[1] World Bank. https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS
[2] GapMinder. https://www.gapminder.org/data/
[3] World Bank. https://data.worldbank.org/indicator/NY.GDP.PCAP.KD
[4] OECD. https://data.worldbank.org/indicator/NY.GDP.PCAP.KD)