Problem Set 3
Statistics, Computation and Applications

Felipe del Canto

October, 2021

Problem 3.1: The Manua Loa CO_2 concentration

For part (a), the linear fit, its residuals and the ACF of the latter, are presented in Figure 1. As can be seen from Figure 1a, the linear model finds a upward trend ($\hat{\alpha}_2 = 0.127$), consistent with the data. However, it does not seem adequate for modeling the trend in the data. The residuals plotted in Figure 1b confirm it. The image shows a strong U-shaped trend left in the residuals that is not accounted for in the linear model. In order to measure the goodness-of-fit of this model, in Figure 1c is presented the autocorrelation function (ACF) of the residuals up to 40 lags. The image shows a positive and statistically significant correlation for the residuals with long-term dependencies. If the model was correctly specified, then we would expect at most a seasonality behavior that in the ACF would show as long-term dependencies and a "symmetric" plot with respect to 0.

For part (b), I repeat the same experiment as in part (a), but with a quadratic model. The corresponding results are presented in Figure 2. In this case, the fit diminishes the coefficient of t ($\hat{\beta}_2 = 0.065$), but includes a small convexity ($\hat{\beta}_3 = 8.692 \cdot 10^{-5}$). The residual plot in Figure 2b shows that the previous U-shaped trend present in the linear fit has dissapeared. However, there still persists an anomaly around 1990 that might correspond to a residual trend not captured by the model. In terms of the goodness-of-fit, the ACF in Figure 2c shows that now the dependencies between residuals are much shorter and the graph is mostly symmetric around 0. This suggests this model is a much better fit than the linear one.

For part (c), I repeat the previous experiments using a quartic model. The corresponding results are presented in Figure 3. The differences of this model with respect to the quadratic one are small. However, some features are salient. For example, in Figure 3b it can be seen that the small variations across time are almost eliminated. What is more, the ACF in Figure 3c exhibits a slightly more symmetric shape, with negative correlations closer to -1, showing that the trend is better captured with this model. Nevertheless, this improvements are really small and the coefficients associated with t^3 and t^4 have orders of magnitude too small (10^{-7} and 10^{-10}, respectively) to consider this model a much better fit than the quadratic one. In general terms, a higher order model will always be able to improve the fit with respect to a lower order one. This, since models with order n are particular cases of models of order $m > n$, where the coefficients for t^k with $k > n$ have been set to 0. However, since the objective is to capture a trend in the data and not fit the data itself, then I would consider increasing the order of the model until there is no apparent trend in the residuals. The latter can be detected visually by plotting the residuals, or through an ACF.

For part (d), the monthly average of the residuals in Figure 2b is presented in Figure 4. The picture shows a strong seasonal component that varies month by month. In Figure 5 is presented the time series with this seasonality extracted, compared with the provided season-adjusted series. From the image, there is no apparent difference between both corrections.

For part (e), in Figure 6a is presented the quadratic fit with the seasonality in Figure 4 incorporated. In general terms, the CO_2 concentration shows an upward trend with a monthly seasonal behavior. The upward trend in consistent with the increase in pollutants that we are observing around the globe and the environmental conse-

Figure 1: Linear trend fit, residuals and ACF of the residuals. The model was fitted using OLS.

(a) Linear fit

(b) Residuals

(c) ACF of residuals
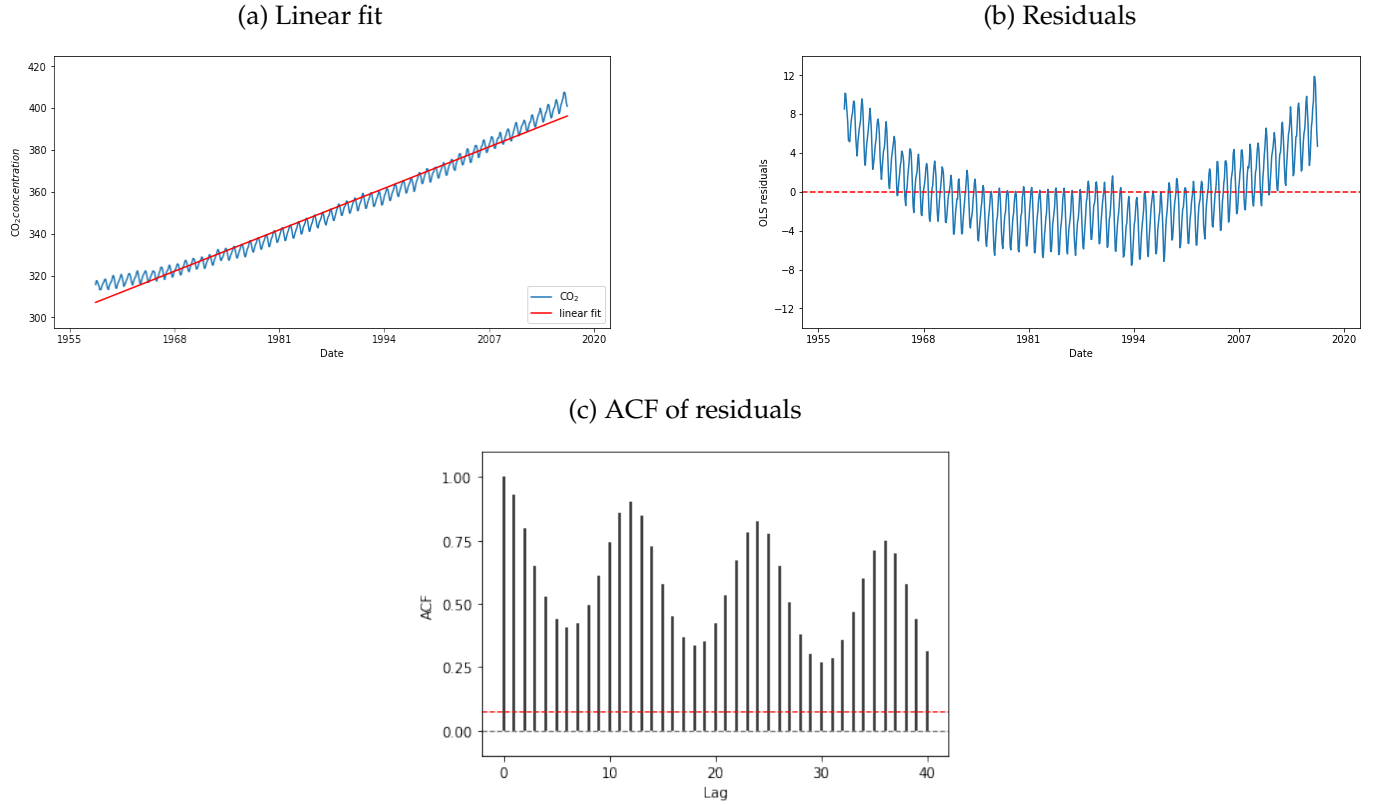
quences therein. In particular, the global warming phenomenon is consistent with an increase in the concentrations of CO_2 in the atmosphere. On the other hand, the seasonality observed may be correlated with the temperature, which changes depending on the season of the year. Under this argument, lower temperatures correlate with higher concentrations of CO_2, since it is more difficult for this gas to escape the atmosphere when temperatures are low (low kinetic energy). On the other hand, when temperatures start to rise, more CO_2 is capable of escaping the atmosphere and consequently, the concentration is lower. Finally, in Figure 6b are presented the residuals of this quadratic fit with seasonal component. The resulting series looks more or less stationary. However, the ACF of the residuals of this series (see Figure 6c) is not consistent with a stationary series, where long-term correlations should fade.

Problem 3.2: PPP Data Analysis

For this problem, the first step is to construct the database. After loading the data, the first step is to keep the relevant observations. In this case, for both the CPI and PriceStats variables, the relevant values will be those of the first of each month. Additionally, the monthly averages of both variables will be stored. After this, the log of all 4 variables is computed. This, in order for these variables to have the same units as the BER rates.

After this, the next step is to merge this dataset with the ones that include the BER for 5 and 10 years. In this case also the first month value and monthly averages are stored. Consequently, after these steps, the final dataset contains 8 variables.

For part (a), since the idea is to predict the values of the CPI (in order to predict inflation), the first step is to choose an appropriate model. For this data generating process, the best option is an AR model. The reason behind this is that prices have a persistent property, in the sense that prices at time t are good reflections for prices in $t + 1$, since consumers base their decisions in those prices. Instead, an MA model, for example, will have some persistency but this is not structural to the model. Consider an MA(1) model. In this case, $X_t = W_t + \theta W_{t-1}$ and, consequently, X_{t+1} will also depend on W_t. However, the main rationale is that prices are correlated

Figure 2: Quadratic trend fit and its residuals. The model was fitted using OLS.

(a) Quadratic fit

(b) Residuals

(c) ACF of residuals
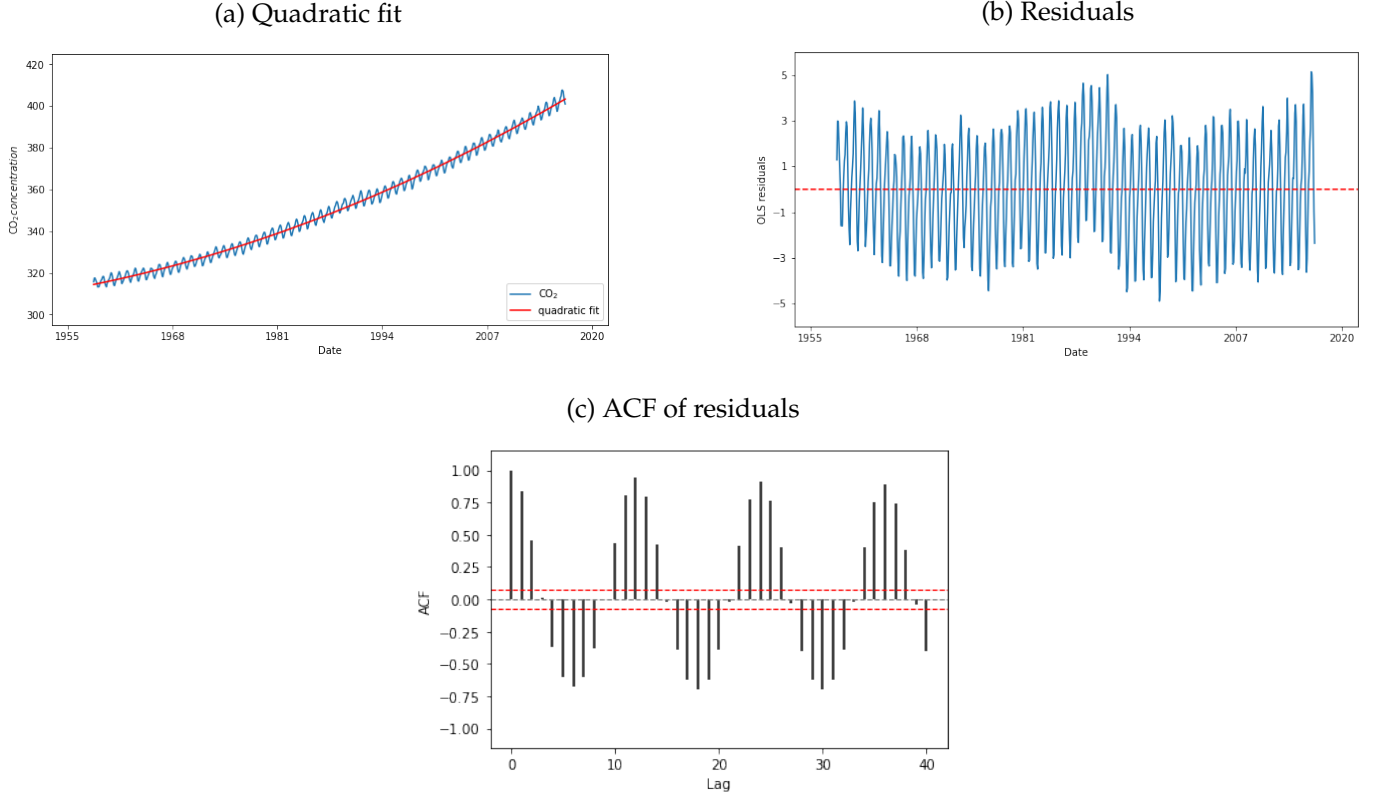
with each other and not only trough with their underlying generation process.

Now, in order to determine the best AR model, the PACF plot is useful. In Figure 7, it is possible to observe a sharp decline in the PACF after lag 1. In fact, from lags 2 to 20, the estimates fall within the 95% confidence interval around 0. Consequently, the image suggests that an AR(1) model may be a good fit for this data. The point estimate for the lagged term in this model is very close to 1 ($\hat{\phi}_1 = 0.9983$) and for this reason the predictions of this model look very similar to the original series, lagged by one month (see Figure 8). The MSPE for this model is $1.571 \cdot 10^{-5}$. It is interesting to note that, even if the AR(1) model is suggested by the PACF, this is not the one that incurs in the lowest prediction error. In that regard, testing the predictions of AR(k) models with $1 \leqslant k \leqslant 10$ resulted in a minimum MSPE of $1.038 \cdot 10^{-5}$, when $k = 3$. Most notably, the MSPE is not monotone in the order of the model.

For part (b), in order to estimate the inflation rate, all the variables in the dataset are used. Based on the AR(1) predictions, the inflation rate in month t, π_t is computed as follows:

$$\pi_t = \left(\frac{\widehat{CPI}_t}{CPI_{t-1}} - 1 \right) \cdot 100, \qquad (1)$$

where $\widehat{CPI}_t$ is the estimated CPI in month t, using the previous AR(1) model. Using the PriceStats data, the inflation rate is computed as:

$$\pi_t = \left(\frac{\text{PriceStats}_t}{\text{PriceStats}_{t-1}} - 1 \right) \cdot 100$$

Finally, since the BER rate is a proxy for the annual inflation rate, the monthly inflation rate is obtained by dividing the BER by 12. This is done for both the first-day value and the monthly average. The inflation estimates are shown in Figure 9. As expected, the CPI forecasts predicted inflation is close to zero. This, because the AR(1) model sets the prediction for the next month almost equal to the previous one (and thus the fraction in (1) is almost 1. Interestingly, the PriceStats predicted inflation is actually very close to the actual inflation. This is expected, since the variable is not being forecasted.

On the other hand, the BER prediction power its low, mostly because this rate is concerned with

Figure 3: Quartic trend fit and its residuals. The model was fitted using OLS.

(a) Quartic fit

(b) Residuals

(c) ACF of residuals
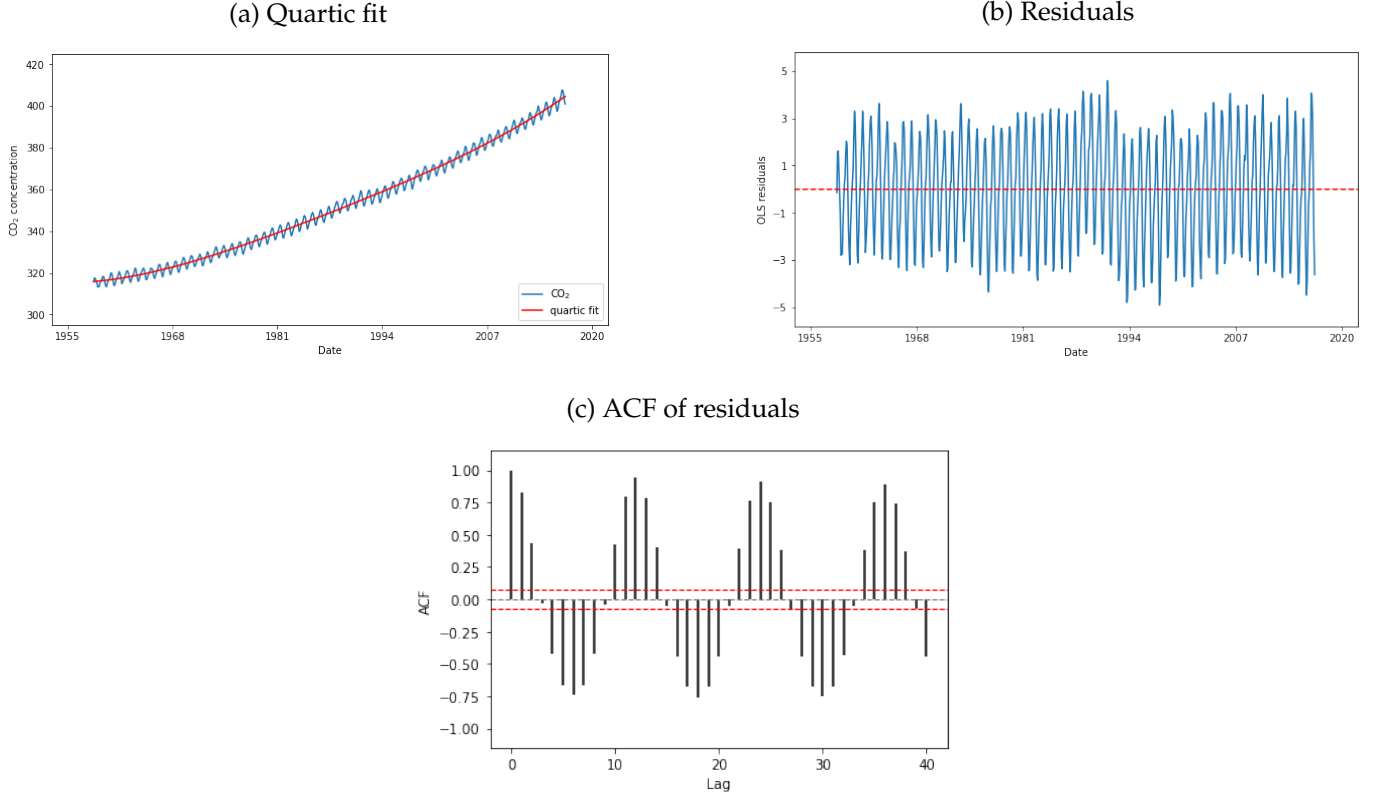
future (average) inflation rates and thus is much smoother than the actual inflation rate. However, it codes expectations (which are crucial to price determination) and thus will be helpful in obtaining better estimates for part (c).

For part (c), to improve the inflation predictions, the AR(1) model in (b) will include exogenous regressors. The idea is that, if the model accurately predicts the next month CPI, then it accurately predicts the inflation rate (see 1)In particular, the data from PriceStats and the BER will be included. This will be done in two ways: using the first-day values and the monthly averages. The predicted series and MSPE are presented in Figure 10 and Table 1, panel B, respectively. The inclusion of these regressors reduces the prediction error in one order of magnitude. It is worth noting that using the monthly averages (instead of the first values of each month) also improves the first prediction (with a MSPE of $6.63 \cdot 10^{-6}$) but it is not as good as using the first values of each month. The reason behind this behavior might be that monthly averages tend to smooth a phenomenon that is highly noisy. Consequently, inflation rates obtained by averaging tend to have a higher prediction error.

For part (d), the same AR(1) model with first-day values used in (c), but including (i) seasonal corrections, (ii) a linear trend, and (iii) both, where fitted. The results are shown in Table 1, panel C. It is interesting to note that none of the three models considered are capable of improving the first-day AR(1) model of panel B. This may be due to deterministic trends having to much incidence in the prediction, meaning less flexibility.

For part (e), to compute the ACF, note that if $h \geqslant 2$, then X_t and X_{t+h} have no W_τ terms in common. Since $\{W_t\}$ are independent, then this implies

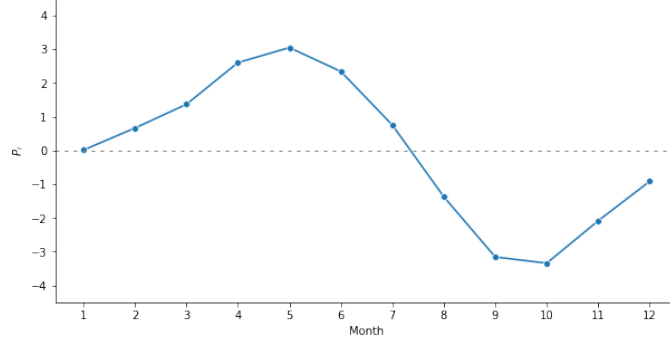$$\gamma(h) := \text{Cov}(X_{t+h}, X_t) = 0, \qquad h \geqslant 2.$$

Thus, the only relevant covariances to compute are the ones with $h = 1$ and $h = 0$. For $h = 1$, we have

$$\gamma(1) = \text{Cov}(W_{t+1} + \theta W_t, W_t + \theta W_{t-1}).$$

Using the fact that the W_t are independent, then

$$\gamma(1) = \theta \text{Cov}(W_t, W_t) = \theta \sigma^2.$$

4

Figure 4: Monthly average of residuals of quadratic model.

Finally, for $h = 0$,

$$\gamma(0) = \text{Cov}(W_t + \theta W_{t-1}, W_t + \theta W_{t-1})$$
$$= (1 + \theta^2)\sigma^2$$

In terms of the mean for this model, we have

$$\mathbb{E}(X_t) = \mathbb{E}(W_t + \theta W_{t-1}).$$

And since W_t, W_{t+1} are independent and have zero mean, then

$$\mathbb{E}(X_t) = 0$$

Altogether, this implies that a time series under a MA(1) model is a zero-mean series whose ACF at lag 0 is equal to $(1 + \theta^2)\sigma^2$ and at lag 1 is equal to $\theta\sigma^2$. At lags $h \geqslant 2$, the ACF should be zero. In particular, if $\theta < 0$, then the ACF is positive at lag 0 and negative at lag 1.

For part (f), let $h > 0$ and $\gamma(h) = \text{Cov}(X_{t+h}, X_t)$, which is independent of t since X_t is stationary. We have that

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t)$$
$$= \text{Cov}(\phi X_{t+h-1} + W_{t+h}, X_t)$$
$$= \phi\gamma(h-1) + \text{Cov}(W_{t+h}, X_t)$$

Since X_t is independent from W_{t+h} for $h > 0$, then

$$\gamma(h) = \phi\gamma(h-1)$$

And thus

$$\gamma(h) = \phi^h\gamma(0)$$

In order to obtain $\gamma(0)$, note that,

$$\gamma(0) = \text{Cov}(X_t, X_t)$$
$$= \text{Cov}(\phi X_{t-1} + W_t, \phi X_{t-1} + W_t)$$
$$= \phi^2\gamma(0) + \sigma^2$$

And thus

$$\gamma(0) = \frac{\sigma^2}{1 - \phi^2}$$

Which implies that

$$\gamma(h) = \frac{\phi^h}{1 - \phi^2}\sigma^2$$

Finally, let $\mu := \mathbb{E}(X_t)$. Using stationarity, we have that

$$\mu = \mathbb{E}(X_t)$$
$$= \mathbb{E}(\phi X_{t-1} + W_t)$$
$$= \phi\mu$$

Figure 5: Deseasoned Manua Loa series. In red, the adjustment was made using the monthly average of the quartic fit residuals. In blue, the provided adjustment.
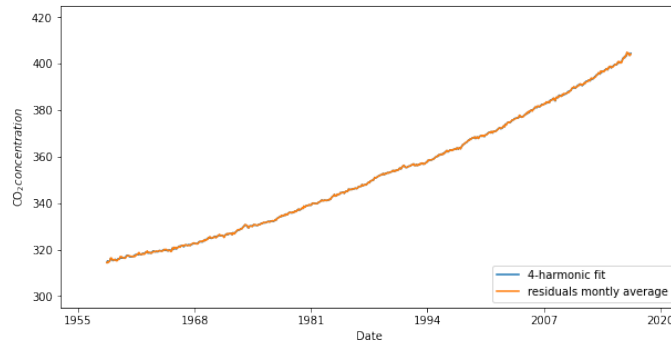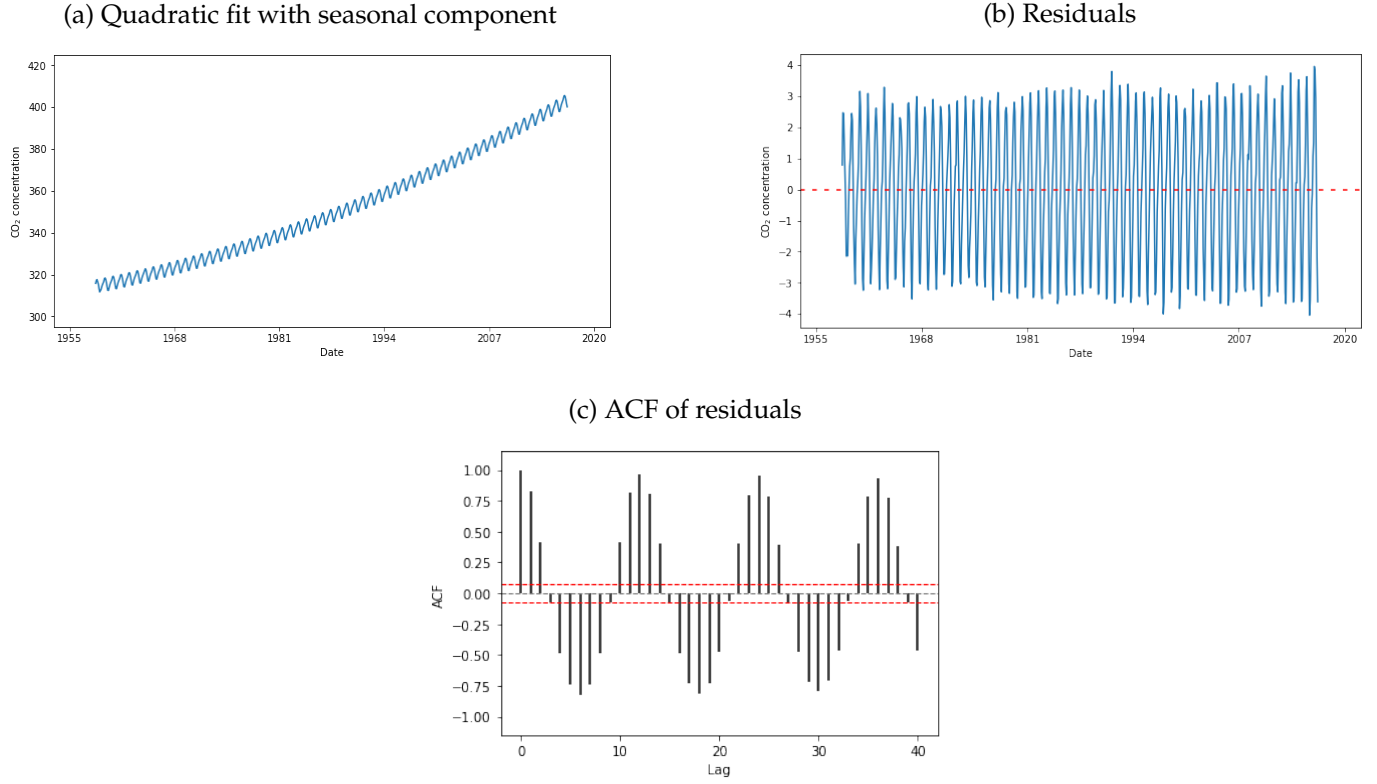

5

Figure 6: Quadratic fit with seasonal component, its residuals, and the ACF of the residuals.

(a) Quadratic fit with seasonal component

(b) Residuals

(c) ACF of residuals

And thus,

$$(1 - \phi)\mu = 0$$

Since $|\phi| < 1$, then $\mu = 0$. In particular, the AR(1) model has the same mean as any MA(q) model. However, note that for the AR(1) model, $\gamma(h) \neq 0$ for all h, whereas an MA(q) model has $\gamma(h) = 0$ for $h > q$. To fix this, consider the following MA(∞) model

$$X_t = \sum_{s=0}^{\infty} \theta^s W_{t-s},$$

where $|\theta| < 1$, and $W_t \overset{iid}{\sim} WN(0, \sigma^2)$. Observe that,

$$X_t = W_t + \theta \sum_{s=0}^{\infty} \theta^s W_{(t-1)-s}$$

$$= \theta X_{t-1} + W_t$$

which means that this model is an AR(1) model. This proves that an AR(1) model is the limit when $q \to \infty$ of a MA(q) model.

6

Table 1: MSPE for AR(1) prediction models. The dependent variable is the log CPI and the exogenous regressors are the log of PriceStats, and the 5 and 10 years BER. These last variables are considered with their first-day value and the monthly average. All values are rounded to the nearest hundredth.

Exogenous regressors	Type	Linear trend	Seasonal Trend	MSPE
Panel A: AR(1) (base specification)				
✗	✗	✗	✗	$1.57 \cdot 10^{-5}$
Panel B: AR(1) with exogenous regressors				
✓	First-day	✗	✗	$6.20 \cdot 10^{-6}$
✓	Monthly average	✗	✗	$6.63 \cdot 10^{-6}$
Panel C: AR(1) with exogenous regressors and trends				
✓	First-day	✓	✗	$7.48 \cdot 10^{-6}$
✓	First-day	✗	✓	$6.40 \cdot 10^{-6}$
✓	First-day	✓	✓	$8.01 \cdot 10^{-6}$

Figure 7: Partial auto-correlation function plot for the log of the Consumer Price Index. The red lines correspond to the 95% confidence interval around 0.
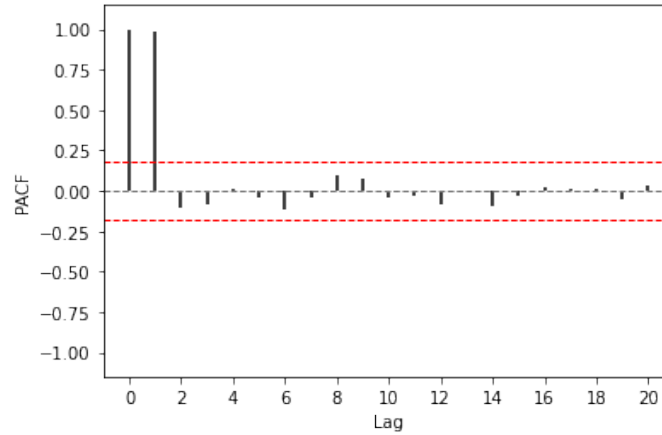

Figure 8: log CPI (in blue) and AR(1) model fit (in red). Since the point estimate for the model is 0.9983, the predicted series looks similar to the original, lagged by one term.
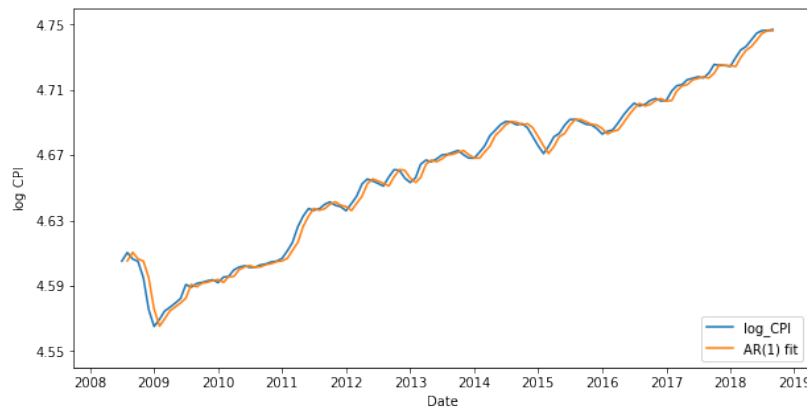
Figure 9: Inflation rate predictions using the AR(1) fit, the PriceStats data, and the 10 and 5 year maturity BER. For this last variable, the first-day value and monthly averages are used.

Figure 10: log CPI (in blue) and AR(1) model fits. The AR(1) model includes the exogenous regressors log PriceStats, and the 5 and 10 year BER. For these last two variables, the first-day value (in red) and monthly averages (in green) are considered.