Forecasting Inflation

Quantitative and Statistical Methods II: Finance

David Catalán

Álvaro Corrales

Jordi Gutiérrez

Manel Pardo

Professor Christian Brownlees

April 1, 2018

Abstract

With more than 50 years of monthly data for the US, we apply forecasting techniques to the growth rates of inflation, including several ARMA models and different linear specifications based on the literature, in addition to ARCH-type models for its volatility. Linear models suggested in the literature and several ARMA models fail to outperform a benchmark out of sample. We also show that ARCH(3) works better in capturing volatility clustering than GARCH(1,1). Nonetheless, the standardized residuals still remain serially correlated and not normally distributed.

1 Introduction

In Economics, inflation and GDP are amongst the most important indicators that policymakers and investors follow. This paper focuses on inflation, a time-series that is key in determining the monetary policy that central banks implement through changes in monetary base.

In 1958, A. W. Phillips realized that unemployment and inflation experienced a negative correlation (using data from the United Kingdom for the period 1861-1957). Policymakers started to implement models based on this observation, which yielded relatively good results up until 70's, at the oil-crisis time, in which the unemployment-inflation correlation became positive. Today, it has been well documented that the relation between inflation and other activity variables is not stable. [1]

The literature identifies several periods for which the relation between activity variables and inflation is notably different. In particular, Kiley (2015) distinguishes between the period from 1976-1995, which witnessed a strong correlation between unemployment and inflation, and the period 1995-2014, where this relation slows down substantially. While several explanations have been proposed (for instance, higher nominal rigidities in periods of low inflation, by Ball and Mazumber (2011), or a clearer commitment to price stability by central banks, by Boivin et al. (2010), among others), the evidence is still inconclusive.

Many authors remain especially puzzled by the unprecedented behavior of the inflation rate during and after the Great Recession. As shown by Blanchard (2016), the relation between unemployment and inflation in the US in the last years is at its lowest. This has important implications for policy-making, as it generates an unprecedentedly strong incentive on the Fed to keep interest rates low, as in principle they could boost employment without worrying too much about inflation.

In any case, the volatile relation between macroeconomic indicators and inflation imposes serious constrains on the forecast on the latter. The immediate consequence is that the accuracy of out-of-sample predictions of this kind of models largely depends on the sample

[1] See, for instance, Stock and Watson (2010)

time frame. Given the limitations of what we could call *Phillips Curve approach*, Stock and Watson (2010) show that the best predictors of inflation tend to be univariate models of the ARIMA family and, especially, variations of the GARCH family.

Following the findings of the literature, the aim of this paper is to test whether or not it is possible to predict inflation, using the latest available data. To do so, we will try to forecast inflation in the US for the period 1964-2017, using several models suggested in the literature. More specifically, we will test the performance of a Phillips-Curve like regression, an ARMA process and a ARCH-type process against the benchmark (Unconditional Mean). In this sense, we will split the sample in 80%-20% such that we use our parameter estimates (in-sample) to evaluate the forecast (out-of-sample). We will do both 1-step ahead and k-step ahead forecast. Interestingly, our remaining 20 percent of the data corresponds to the period of the Great Recession and afterwards, so we can expect structural differences between our two samples.

To define the Phillips Curve, several indexes will be tested: the industrial production index, the unemployment level, average hourly wage, the price of imported goods index and the consumers confidence index. Other controls include money supply and the Fed's policy rate.

2 Description of inflation time-series

In Figure 1 we show the plot of the CPI. We can observe that there's a positive trend, indicating us that prices are increasing over time. However, we need a stationary process so that we can analyze it and forecast future realizations of such a process.

In Table 1 we show the first four moments of all our variables. It is important to note that our dependent variable, Inflation, is on an index scale with base $100 = 1982 - 84$. This implies that its mean is not directly interpretable, as the index, when compared to the base years, has been growing almost uninterruptedly. Other variables, such as wages, Money2 or Industrial production face a similar increasing trend if we just look at the values or indices. In the case of consumer confidence, the mean is close to its base value

100.

In Table 2 we show the p-values of the Augmented Dickey-Fuller, the Jarque-Bera and the Ljung-Box test. Only for the inflation index we could marginally reject the null hypothesis of unit root, but we still turn all variables into growth rates to gain as well on interpretability. In addition, in all cases we reject normality and serial independence.

In Table 3 we show the first moments of the growth rates of our dataset. Inflation has a slightly positive mean growth rate, as do Industrial Production and Wages. We can see, on average, an expansionary monetary policy in the negative and positive means of the policy rate and Money 2 growth rates. In addition, Consumer confidence, the interest rate on businesses and, to a lower degree, unemployment, all have zero mean. Policy rate is by far the variable with the highest variance of our sample, whereas consumer confidence, unemployment and inflation are those with the lowest volatility. The growth rate of wages is the variable with the highest positive skewness while industrial production has the most left-skewed distribution, as there have been economic periods with sharp decreases in economic activity. The high volatility of the policy rate is then translated into it being the variable with the thickest tails. Our main variable, inflation, shows positive skewness and average kurtosis when compared with the rest of the sample.

In Table 4 we show the p-values of the same tests on the growth rates. Now, in all cases we reject the null of stationarity although, as before, none of the variables has a normal distribution. or serial independence, according to the Jarque-Bera and the Ljung-Box test, respectively.

In Figure 2 we show the distribution of our variable of interest (first-differences on CPI or inflation). We can observe that inflation is distributed with large tails (high kurtosis) and is slightly right-tailed (positive skewness). In Figure 3 we can highlight that this is not distributed as a normal, and in Figure 4 we plot the persistent autocorrelation function for two years of data.

3 In-Sample Analysis

We split our sample following a 80%-20% rule such that:

- In-sample: monthly data from 1964-01-01 to 2006-12-31, which are 516 observations.
- Out-of-sample: monthly data from 2007-01-01 to 2017-10-01, which are 130 observations.

3.1 Linear model

Our first attempt in order to analyze inflation is to run 3 different linear model specifications. The first consists in regressing inflation on unemployment alone, that is, the simplest Phillips curve. The second specification (ARX) regresses inflation on unemployment and the other potential predictors commented in Section 2 . The last specification follows closely Blanchard, Cerutti and Summers (2015), where they use a richer Phillips curve specification, including long term inflation expectations and the difference between unemployment and natural unemployment rate. However, we keep the data structure at the monthly level, while Blanchard, Cerutti and Summers use quarterly data.

Adding the above mentioned variables in the third specification substantially reduces the number observations we can use due to data availability. Hence, we finally discard it. Out of the other 2 specifications the one performing better is the second specification (ARX). This is probably due to the fact that it adds more predictors compared to the first specification. Looking at the residuals autocorrelation we see how in the second specification there is less autocorrelation than in the linear model with just unemployment. However, there is still a lot of serial dependence left. By running a Ljung-Box test, we reject the null of no autocorrelation of the residuals at a 1 percent level. We also add a Q-Q plot of the residuals of the first and second specifications against the normal distribution in Figure 7 and the fitted values against the actual values in Figure 8.

Overall, the linear model is a very poor specification to predict inflation. Note that this is consistent with the findings in the literature, as a constantly-changing parameter of the

Phillips curve would imply that our estimator for the whole sample would fail to capture much strutural variation.

3.2 ARMA models

The fact that prices are sticky is a well-known theoretical result in the literature of Monetary Economics [2]. Empirically, some authors have taken different approaches to test this claim: Romer & Romer (2004) have pointed out that changes in monetary policy take several months to be fully reflected by changes in prices. Nakamura & Steinsson (2013) use data on prices from the US to show that median time duration of prices ranges from 7 to 9 months, depending on the date frame, whereas the mean duration can rise to almost 12 months.

These findings have important implications for our study if we are to predict inflation using an ARMA model. In particular, to capture some variation in our measure of inflation it may be necessary to include a substantial amount of monthly lags. Consequently, the parsimoniuos and simplest models such as AR(1), MA(1) or ARMA(1,1), would not perform particularly well in principle. We can easily test this by fitting the models mentioned above and running a Ljung-Box test with 24 lags (two years) on the residuals. In all three cases, we reject the null hypothesis that the residuals do not exhibit dependence (that is, the model captures all of the time variation) with a p-value close to 0.

Considering these results, and having in mind the potential lags implied by the literature, we fit increasingly less parsimonious ARMA models and carry out a Ljung-Box test on the residuals to see at which point all serial dependence is eliminated. The p-values for the null of no dependence for the residuals of several models can be seen in Table 6 of the Appendix. As shown in the literature, there is still a certain degree of price stickiness after 6 and even 9 months.

The p-values of the Ljung-Box tests suggest that we should focus our analysis on the AR(12), ARMA(12,1) and ARMA(9,6) models. The parameter estimates of the models

[2] For more details, see Galí (2008)

are reported in Table 7. Interestingly, and in line with what is predicted in the literature, the most significant parameters in the the AR(12) and the ARMA(12,1) models are those corresponding to lags 9, 11 and 12. In other words, the price level at any given month, t, is heavily determined by prices 9, 11 and 12 months before t. To assess the in-sample goodness-of-fit of the three models, we compute the AIC and BIC, which we complement with the negative log-likelihood (nLL). Table 8 contains these three indicators for the three models. While all three indicators are very similar across models, they all point at the AR(12) as the best model for in sample prediction, followed by the ARMA(12,1). The fitted values of the three models against the actual values can be seen in Figure 10. A Q-Q plot of the residuals of the three models against the Normal distribution is also provided in Figure 11. Graphically, we can already see that, while they are not serially correlated (as shown in Table 6), they are far from being normally distributed.

3.3 ARCH & GARCH

For Central Banks, price stability is a key goal in their policy since it is widely known that inflation volatility is usually related with higher risk premium, hedging costs and unforeseen distribution of wealth[3]. Here, we want to check if periods of high inflation volatility tend to be followed by periods of high inflation volatility. Then, more technically, we want to analyze volatility clustering and introduce a non-linear dynamic model which is able to capture it.

AutoRegressive Conditional Heteroskedasticity (ARCH) model [4] is a non-linear quantitative model that is able to capture volatility clustering and was widely used in financial time series. It uses a deterministic approach on the volatility. Here, we use the most common approach in literature, an ARCH(3)

ARCH(3) model specification:

$$i_t = \mu + \sqrt[2]{\sigma_t^2} z_t \qquad\qquad z_t \sim \mathcal{N}(0,1) \tag{1}$$

[3]Phillip C. Rother (2004)

[4]Robert Engle (1982)

where μ is the inflation mean, i_t is inflation rate and σ_t^2 is the variance, which follows:

$$\sigma_t^2 = \omega + \alpha_1 i_{t-1}^2 + \alpha_2 i_{t-2}^2 + \alpha_3 i_{t-3}^2 \tag{2}$$

where $\omega > 0$ and $0 \leq \sum_{i=1}^{3} \alpha_i < 1$

GARCH is just a parsimonious specification of ARCH. We use an GARCH(1,1) model specification:

$$\sigma_t^2 = \omega + \alpha i_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3}$$

where $\omega > 0$, $0 \leq \alpha + \beta < 1$

First, we use an ARCH-LM test to assess if there's evidence of volatility clustering. After observing p.value = 0.000, we can reject the null of no volatility clustering.

Then, we proceed to compute an ARCH(3) and a GARCH (1,1) model to see if these models are able to capture volatility clustering. We fit the parameter estimates in Table 9. Note that the persistence of parameter estimates is less than 1, which implies that the conditional variance of inflation will converge to the unconditional mean.

Third, we measure goodness-of-fit to see which model features the best in-sample performance. See Table 10. Log-likelihood, AIC and BIC all show that ARCH(3) features better than GARCH(1,1), in-sample.

Fourth, we provide some fitted values for ARCH, GARCH and we compare it with a proxy of inflation volatility (absolute values of inflation). We also provide rolling variance with window equal to 24 (2 years). See Figure 12 for ARCH vs Rolling variance vs Proxy. See Figure 13 for GARCH vs Rolling variance vs Proxy. See Figure 14 for ARCH vs GARCH vs Rolling. In-sample, ARCH(3) and GARCH(1,1) fits really similar.

Finally, we provide some residual diagnostics for ARCH and GARCH. First and foremost, we would like the absolute standarized residuals not to be auto-correlated. In Figure 15 we can see how ARCH is able to capture majority of volatility clustering, but once we apply the Ljung-Box test, we notice that we reject the null of no serial correlation. In Figure 16 we observe how GARCH is also able to capture a high part of volatility clustering, but worse than ARCH since Ljung-Box test sustains a higher value. Thus, we also reject the null of no serial correlation. Now, we would also like that residuals are normally distributed. In Figure 17 we see how standardized residuals in ARCH improves the normality of inflation, but we still reject the null of normality (by the Jarque-Bera test). In Figure 18 we notice that standardized residuals in GARCH improves even more the normality of inflation, but we still reject the null of normality (JB-test).

4 Out-of-Sample Prediction

4.1 Linear Models

We focus on Specification 2 (ARX) for out-of sample analysis. Figure 9 shows the 1-step ahead forecast where, in line with the in-sample analysis, the residuals have a lot of serial correlation and the model does not capture well the future variation of inflation.

In Table 11 we show the MSE (Mean Squared Error) as well as the Diebold Mariano test, the test is not able to reject the null hypothesis that the model's predictive ability is statistically the same as the predictive ability of the unconditional mean. We get a better MSE in the second specification than in the simple linear model with just unemployment.

4.2 ARMA models

Figures 19 and 20 show the 1-step-ahead and dynamic forecasts of each model, respectively. As we would expect, the dynamic forecast fails to capture the future variation of the process, and it reverts to its in-sample unconditional mean (as it is the data that we used to train the model).

On the other hand, we do not appreciate substantial differences in 1-step-ahead predictive ability between the three models at first sight. This intuition is confirmed when we compute the MSE of the out-of-sample forecasts. In all three cases, the MSE is almost identical (even if slightly smaller for the AR(12)), meaning that differences between the actual series and the forecast are negligible (first row of Table 12).

To test their forecasting ability, we run several Diebold-Mariano tests on the null hypothesis that the models' predictive ability is the same as the unconditional mean of the in-sample series. The p-values associated to each test can be seen in the second row of Table 12. In all three cases, while the p-value is not very large, we nonetheless fail to reject the null hypothesis.

4.3 ARCH & GARCH

Now, we use our parameter estimates in-sample in ARCH(3) and GARCH(1,1) to forecast out-of-sample volatility, as shown in Figure 21. Again, we use MSE as the loss function to evaluate the performance of the models, see Table 13.

We can observe that ARCH(3) fits better out-of-sample, in line with the in-sample analysis. However, MSE are not so far away one from each other. Also note that our proxy for volatility (absolute inflation adjusted by mean inflation) is more noisy that our both models. We see that in general, ARCH(3) and GARCH(1,1) behave well in capturing volatility clustering but it is not enough. As studied in the in-sample analysis, standardized residuals are still auto-correlated and not normally distributed.

For future analysis, we recommend to test for asymmetric effects and forecast with models that allow for them (TARCH, EGARCH).

5 Conclusions

In this paper we have tried to forecast inflation, a time series that is stationary, not normal, and which shows serial correlation, using three of the most widely used models in time series analysis. First, we have tested a kind of linear regression models that rely on other variables, such as Unemployment, or wages (*Phillips-Curve* type of models). They do not seem to perform significantly better than the unconditional mean of the series. In particular, these variables do not help in adjusting for the small variations of the data. Second, we have tested some ARMA models that try to account for the price stickiness widely documented in the literature. While these models capture a substantial variation of the series, in general they also fail to perform better than the unconditional mean of the in-sample series. Finally, we have also tested a volatility modeling approach. In this sense, when we model the volatility with an ARCH(3) and a GARCH(1,1), both models are able to capture volatility clustering, but in-sample information criteria and Ljung-Box test give us evidence to choose ARCH(3) instead of GARCH(1,1). In any case, it is important to highlight that the residuals of both models are still serially correlated and they are not normally distributed.

In conclusion, none of the models that we have tested in this paper does a particularly good job at forecasting inflation, even considering a relatively large period of time (from 1964 to 2017). This has important theoretical and practical implications, as existing models are not able to capture and forecast structural changes that may have disproportionately large impacts on inflation and all macroeconomic and financial indicators that depend on it.

Appendix

Figure 1: CPI time series

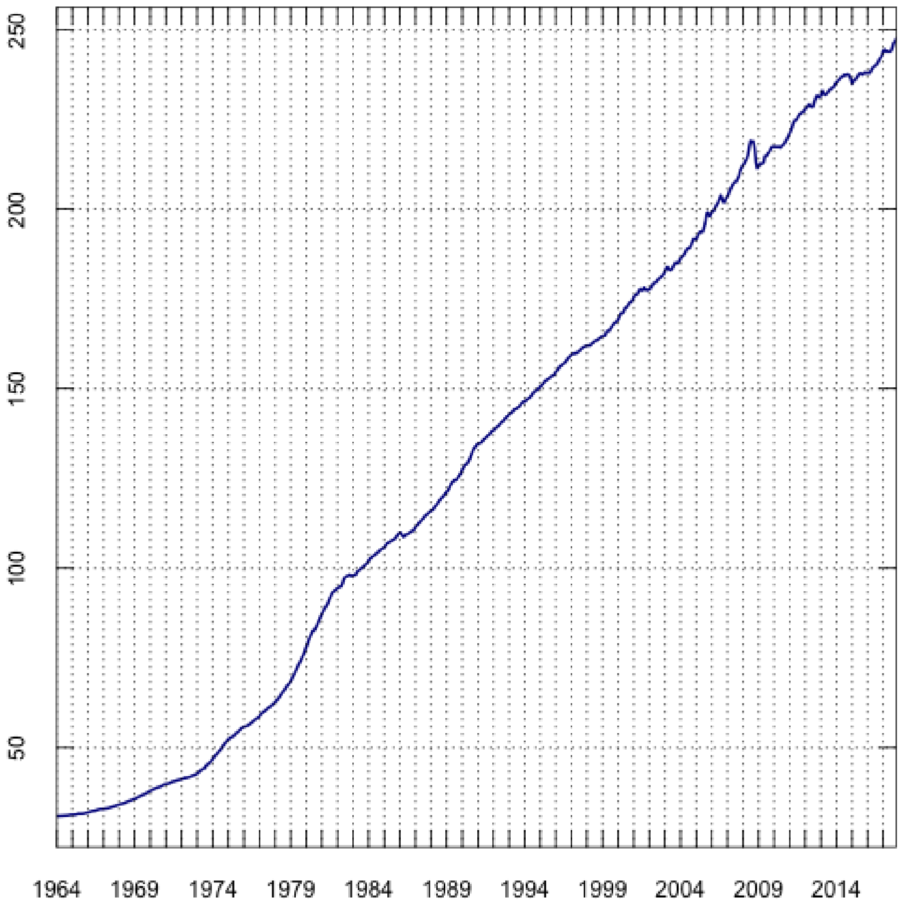

Table 1: Moments

	Indpro	Inflation	Irate B	Irate Fed	M2	Unemp	Wages	Cons Conf
Mean	69.56	130.746	7.425	5.276	4164.494	6.057	10.917	99.926
SD	24.376	69.185	3.375	3.757	3559.899	1.633	5.926	1.422
Skewness	0.099	0.04	1.193	0.787	1.023	0.66	0.258	-0.405
Kurtosis	1.528	1.691	5.157	4.045	3.078	2.851	1.866	2.441

Table 2: P-values

	Indpro	Inflation	Irate B	Irate Fed	M2	Unemp	Wages	Cons Conf
ADF	0.23	0.047	0.11	0.078	0.99	0.084	0.634	0.077
JB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3: Moments of growth rates

	Indpro	Inflation	Irate B	Irate Fed	M2	Unemp	Wages	Cons Conf
Mean	0.203	0.322	-0.009	-0.170	0.548	-0.048	0.338	0.000
SD	0.737	0.32	3.816	10.292	0.604	2.85	0.244	0.209
Skew	-0.989	0.173	-0.352	-2.004	-0.271	0.453	0.928	0.058
Kurt	8.135	7.489	11.799	26.608	3.569	3.991	5.397	4.442

Table 4: P-values of growth rates

	Indpro	Inflation	Irate B	Irate Fed	M2	Unemp	Wages	ConsConf
ADF	0.010	0.010	0.010	0.010	0.010	0.010	0.021	0.010
JB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 2: Distribution of inflation

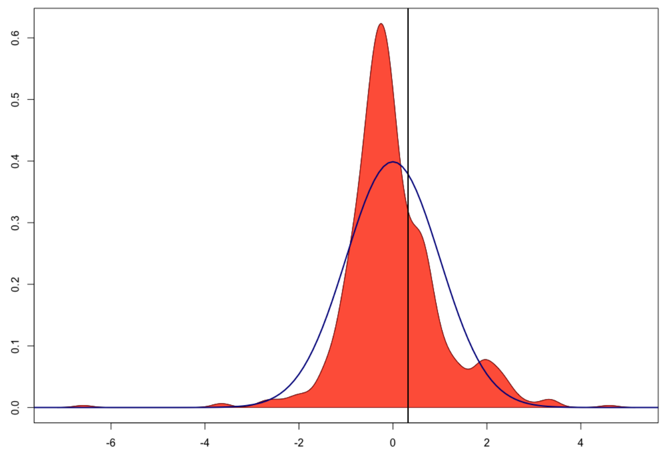

Figure 3: Q-Q Plot of inflation against Normal distribution

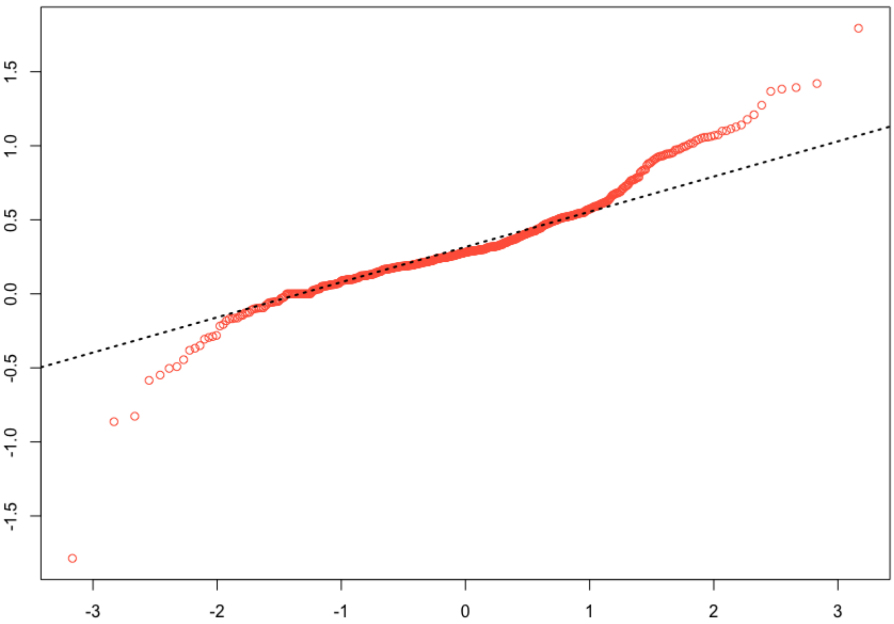

Figure 4: ACF of growth inflation

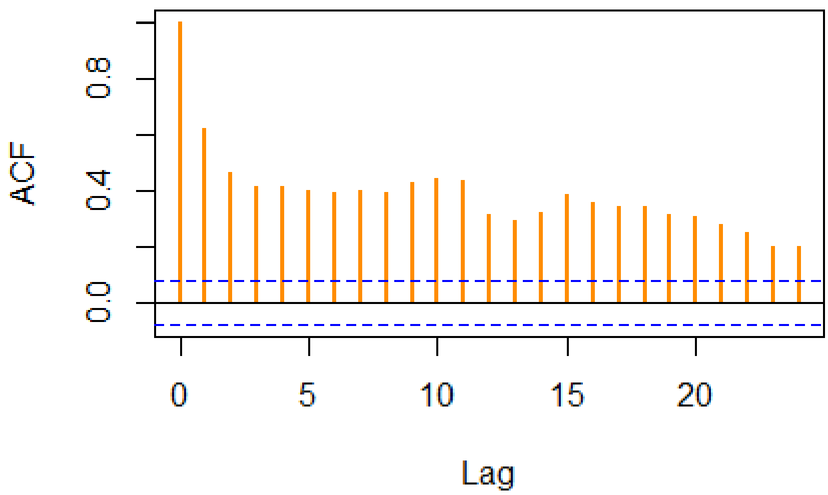

Table 5: ARX. Parameter estimates

Variable	Parameter estimates
(Intercept)	0.199
indpro	-0.059
money	0.005
unemp	0.009
iratefed	0.002
irateb	0.016
wages	0.474
cons_conf	-0.15

Figure 5: Linear. ACF residuals

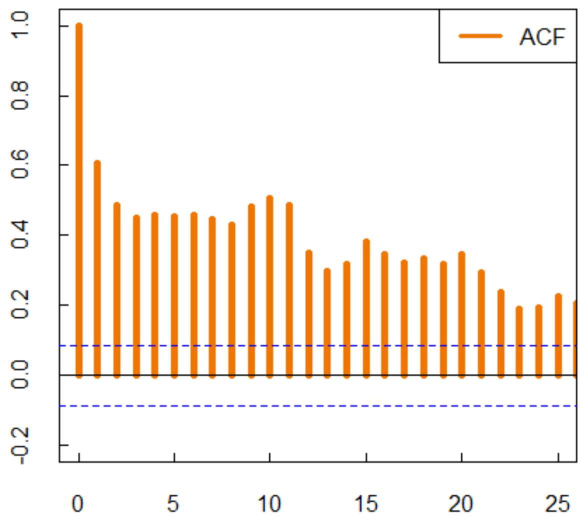

Table 6: ARMA. Ljung-Box tests on residuals (p-values)

	AR(6)	ARMA(6,1)	ARMA(6,6)	AR(9)
LB	0.000	0.000	0.012	0.002

	ARMA(9,1)	ARMA(9,6)	AR(12)	ARMA(12,1)
LB	0.021	0.762	0.794	0.911

Figure 6: ARX. ACF residuals

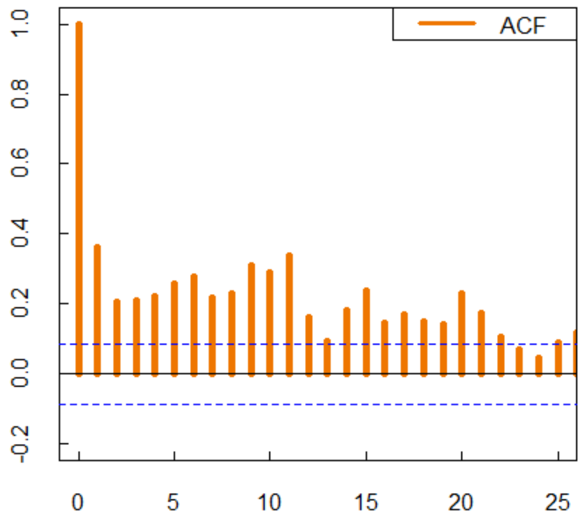

Figure 7: Q-Qplot residuals Linear and ARX

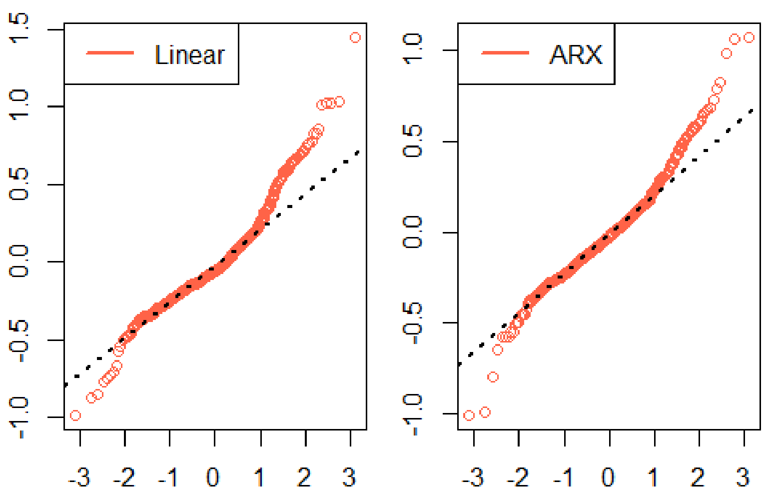

Figure 8: ARX. Fitted values

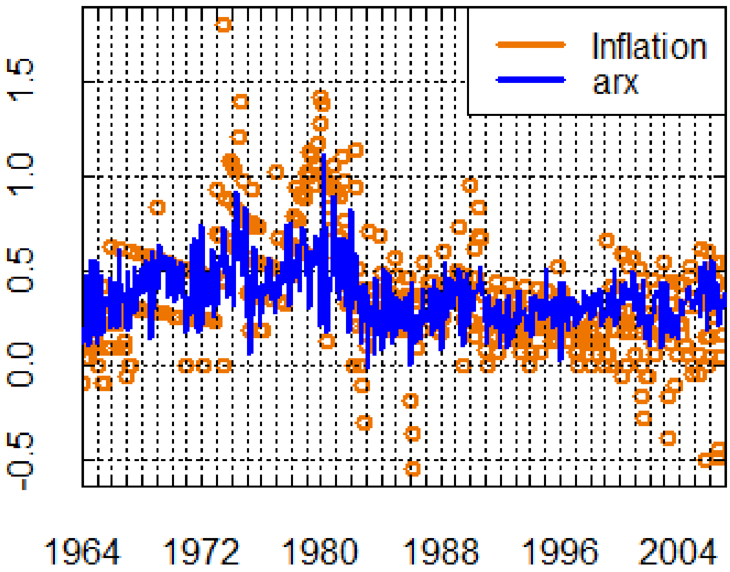

Figure 9: ARX. Prediction

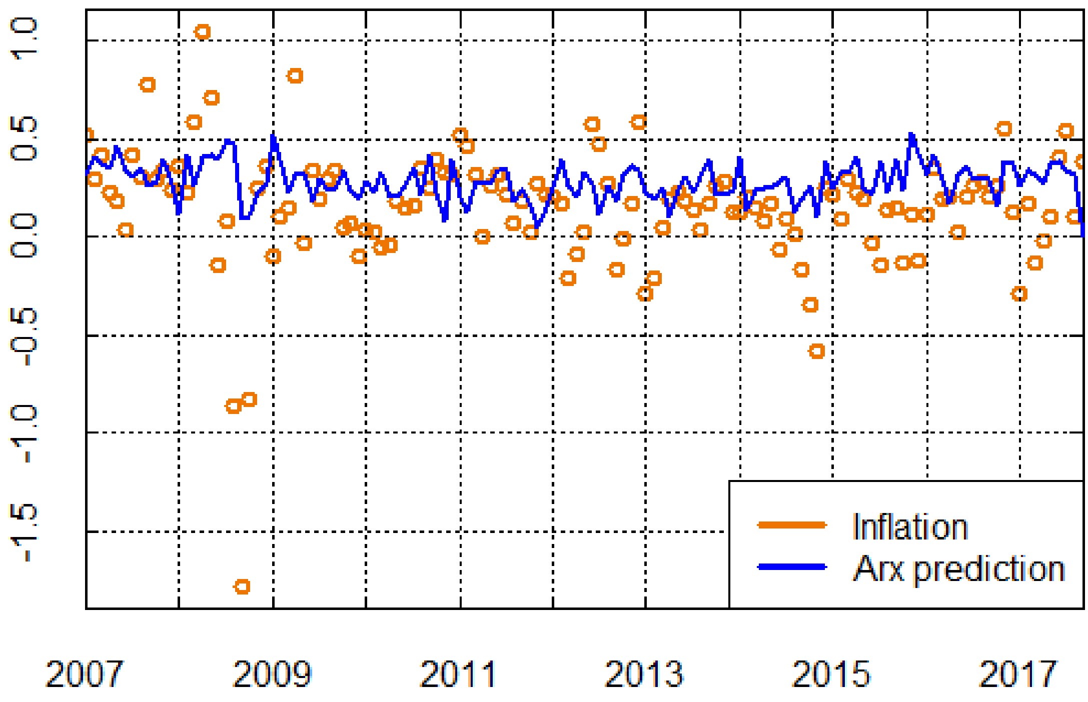

Figure 10: ARMA. Fitted values

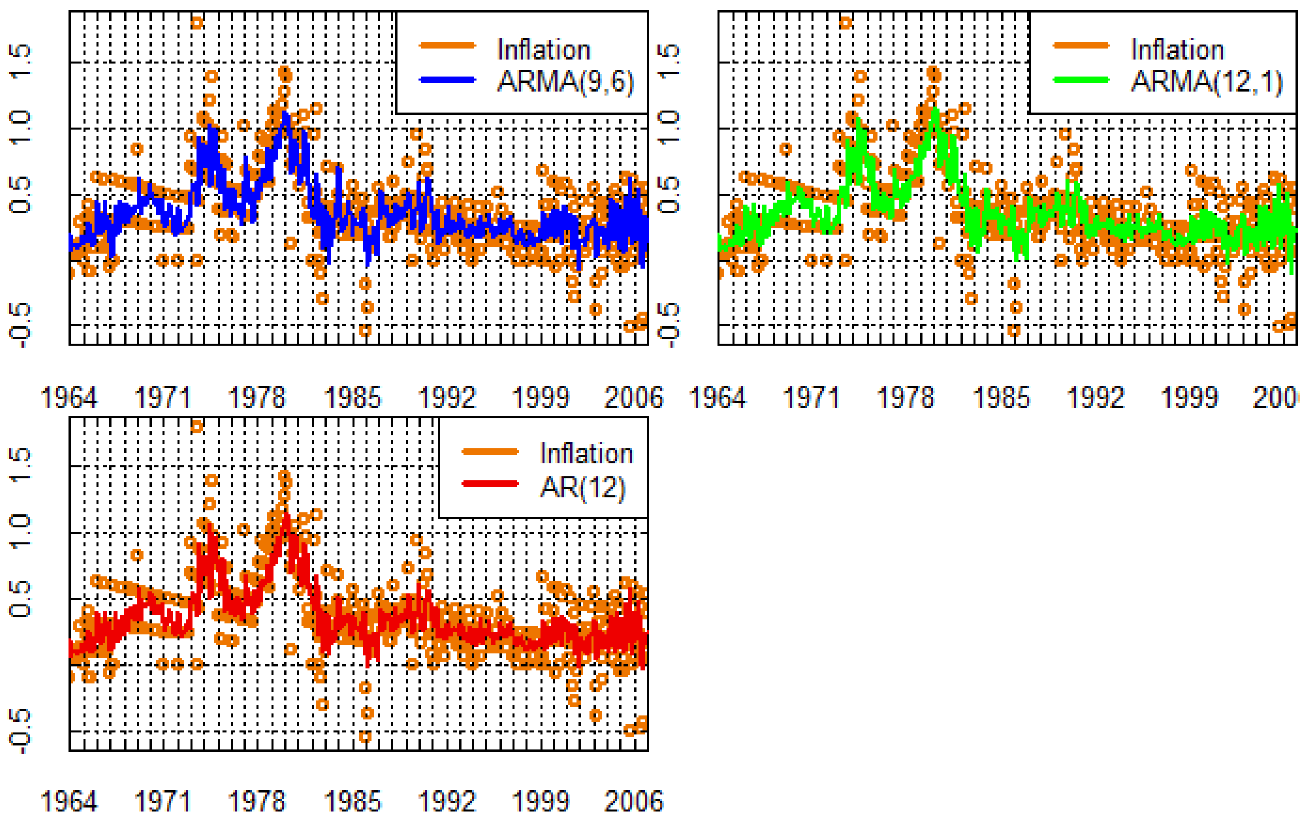

Table 7: ARMA. Parameter estimates

	ARMA(9,6)	AR(12)	ARMA(12,1)
Intercept	0.302***	0.307***	0.307***
ϕ_1	0.658***	0.451	0.443*
ϕ_2	0.331**	0.009	0.012
ϕ_3	-0.890***	0.034	0.034
ϕ_4	0.630***	0.072*	0.072*
ϕ_5	0.469***	0.028	0.028
ϕ_6	-0.548***	0.036	0.036
ϕ_7	0.175**	0.063	0.063
ϕ_8	-0.041	0.012	0.012
ϕ_9	0.143***	0.098**	0.098**
ϕ_{10}		0.061	0.062
ϕ_{11}		0.135***	0.135***
ϕ_{12}		-0.132***	-0.131***
θ_1	-0.206		0.007
θ_2	-0.419***		
θ_3	0.746***		
θ_4	-0.263*		
θ_5	-0.597***		
θ_6	0.353**		

Table 8: ARMA. In-sample Goodness-of-fit

	AR(12)	ARMA(12,1)	ARMA(9,6)
nLL	62.943	65.301	66.214
AIC	-0.189	-0.195	-0.190
BIC	-0.074	-0.071	-0.051

Figure 11: ARMA. Q-Q plot of residuals against Normal distribution

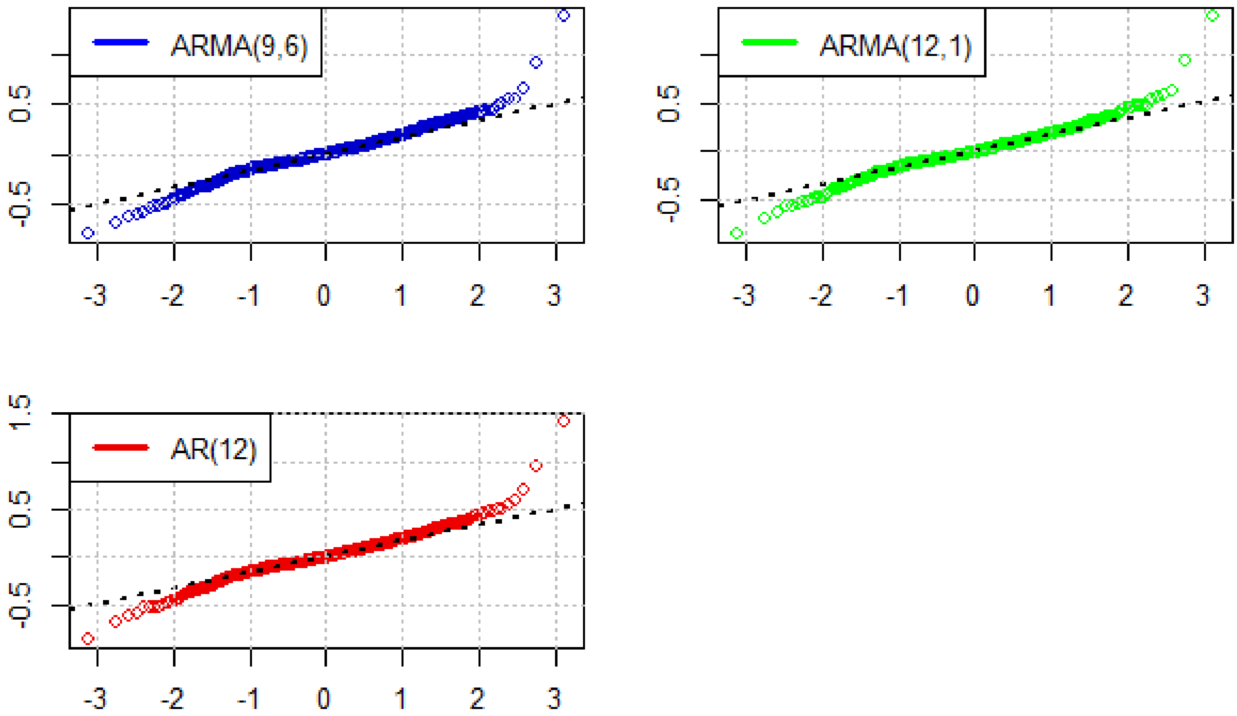

Table 9: ARCH(1) & GARCH(1,1). Parameter estimates

	ω	α_1	α_2	α_3	β
ARCH(3)	0.053	0.330	0.207	0.216	-
GARCH(1,1)	0.006	0.197	-	-	0.775

Table 10: ARCH(1) & GARCH(1,1). Goodness-of-fit

	Loglike	AIC	BIC
ARCH(3)	-204.468	0.812	0.853
GARCH(1,1)	-214.503	0.843	0.868

Figure 12: ARCH. Fitted values

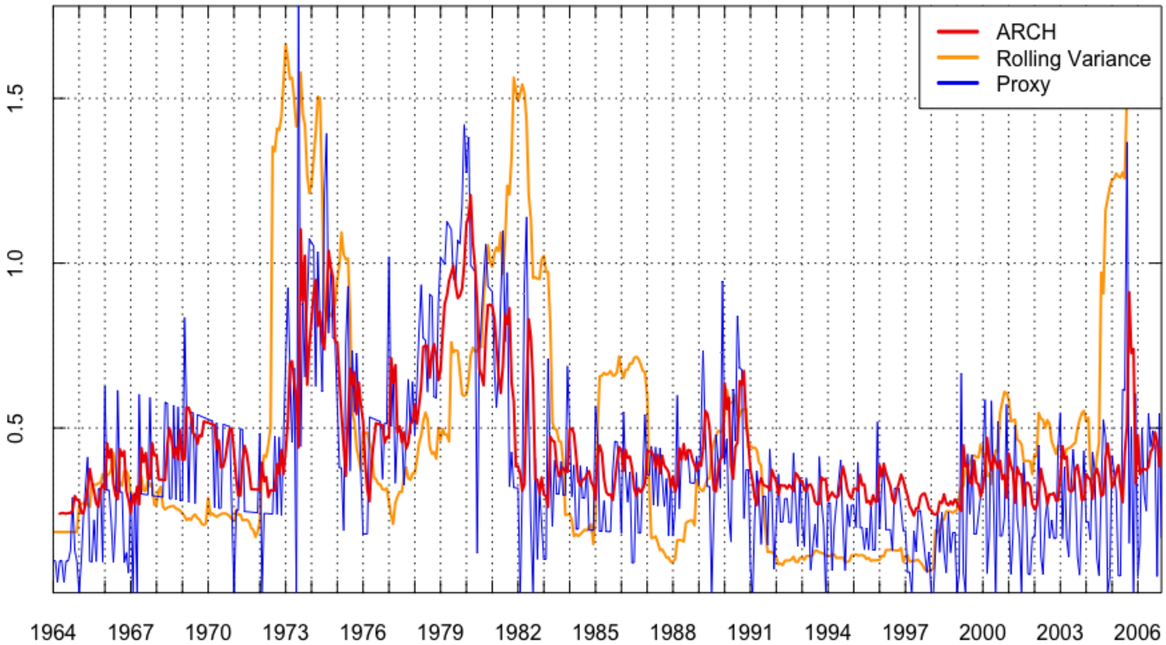

Figure 13: GARCH. Fitted values

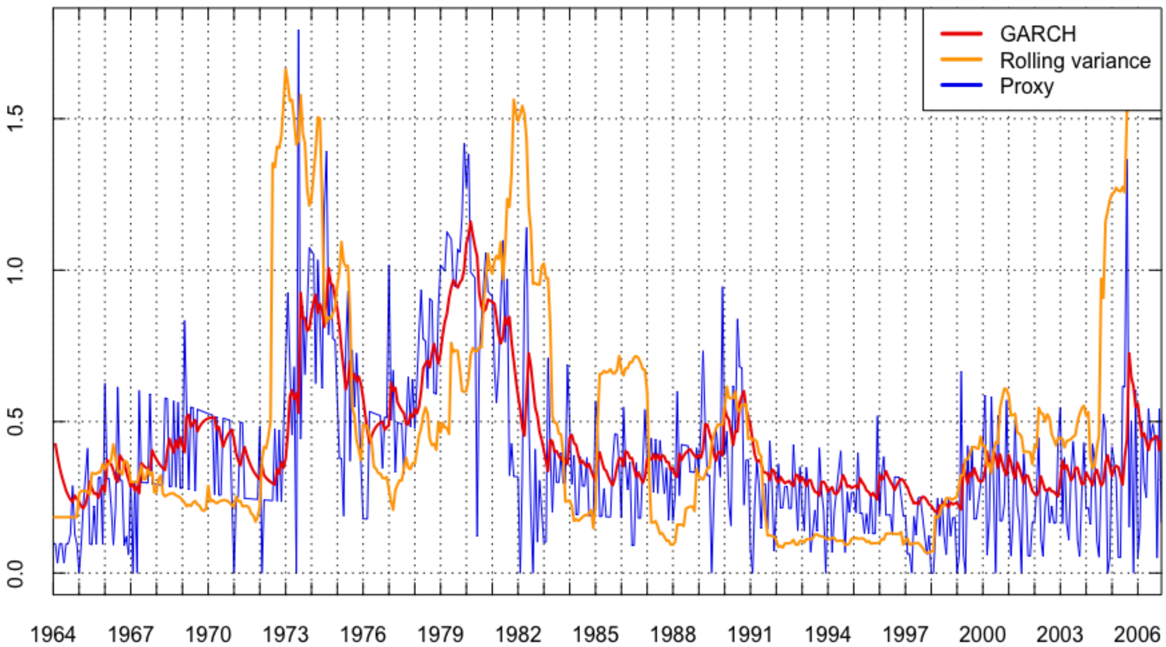

Table 11: Linear and ARX. MSE and DM

	Linear	ARX
MSE	0.146	0.118
DM	0.142	0.17

Figure 14: ARCH vs GARCH. Fitted values

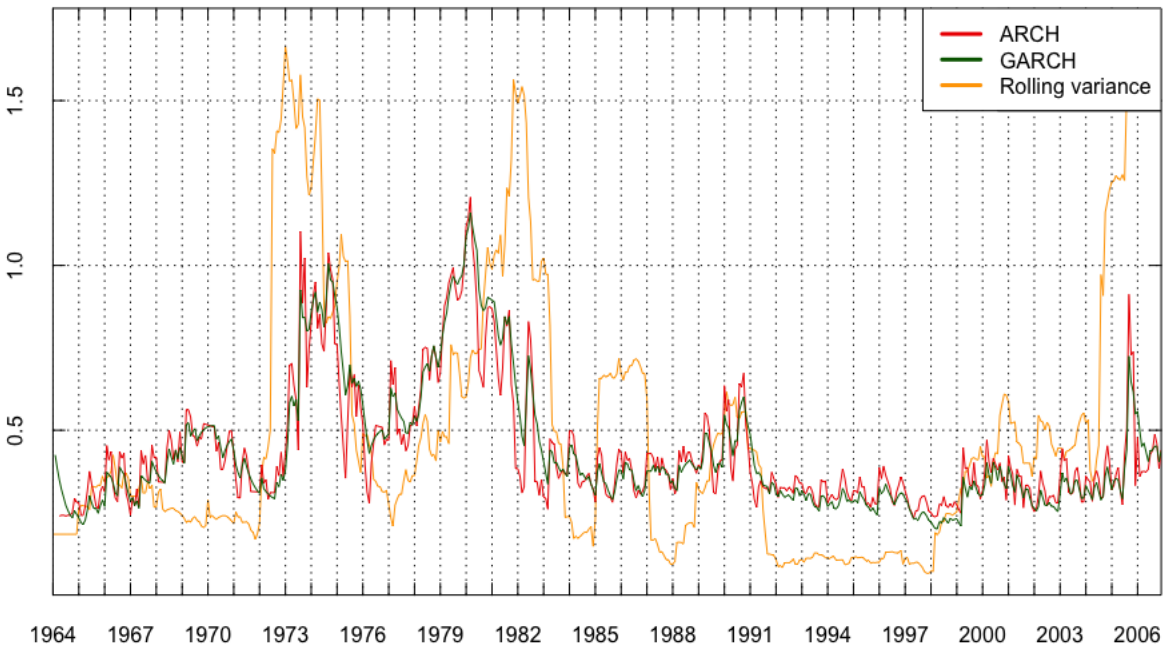

Figure 15: ARCH. ACF and LB test

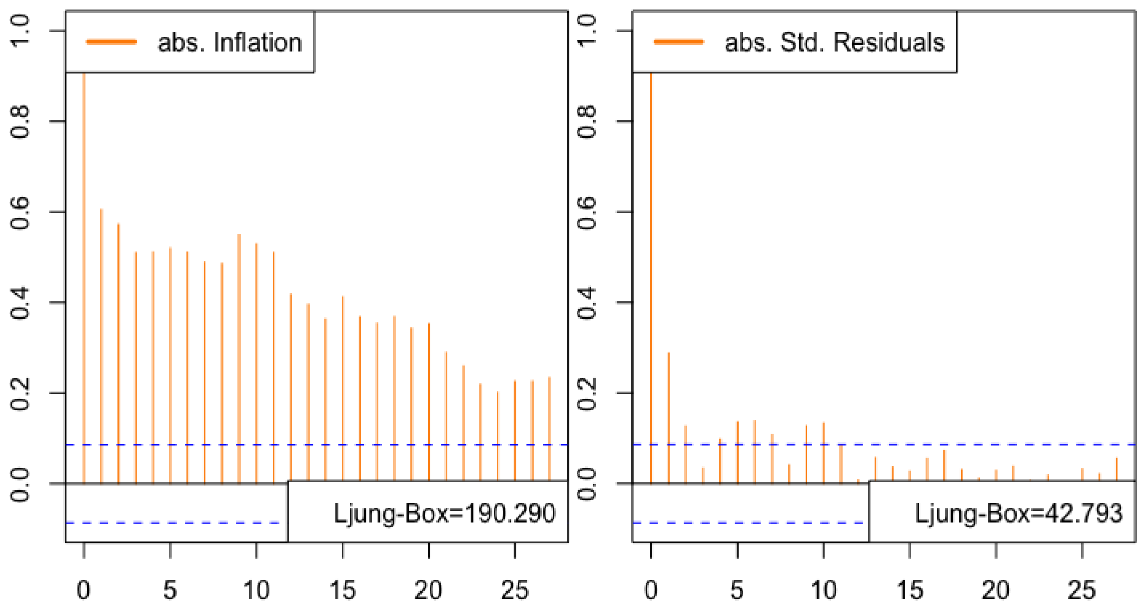

Table 12: ARMA. Out-of-sample MSE and DM

	ARMA(9,6)	AR(12)	ARMA(12,1)
MSE	0.098	0.093	0.095
DM	0.174	0.181	0.177

Figure 16: GARCH. ACF and LB test

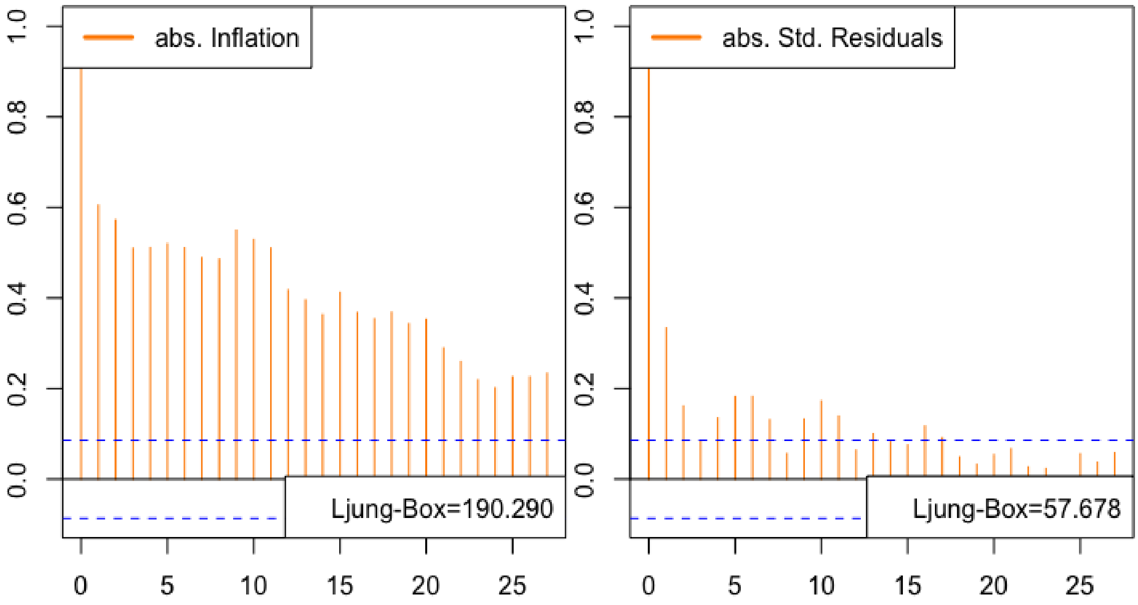

Figure 17: ARCH. Normality diagnostic

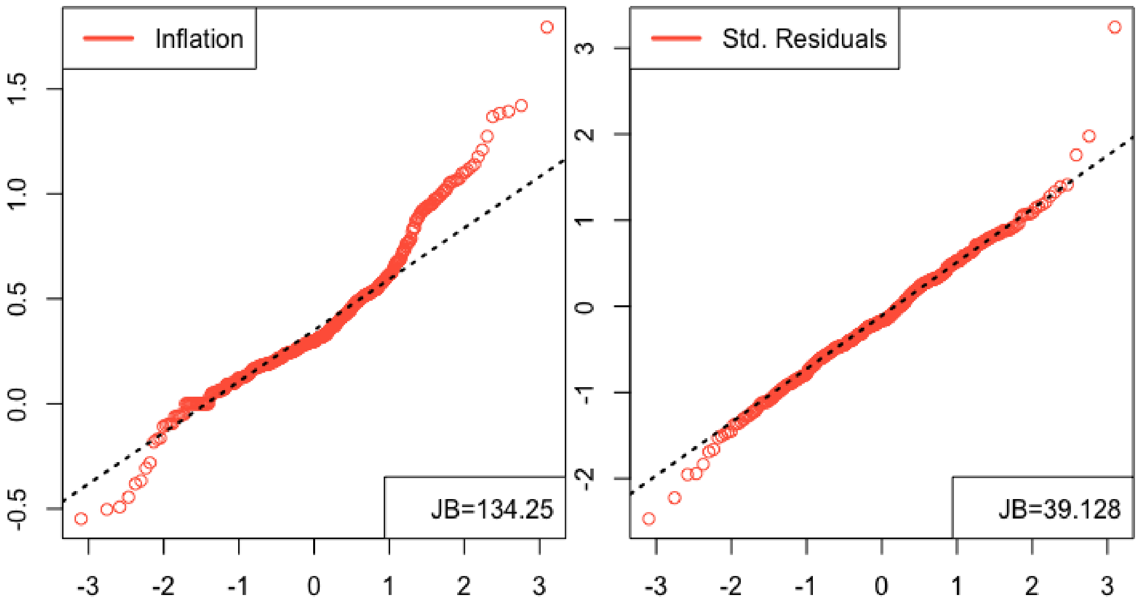

Table 13: ARCH & GARCH. Out-of-sample MSE test

	ARCH(3)	GARCH(1,1)
MSE	0.030	0.038

Figure 18: GARCH. Normality diagnostic

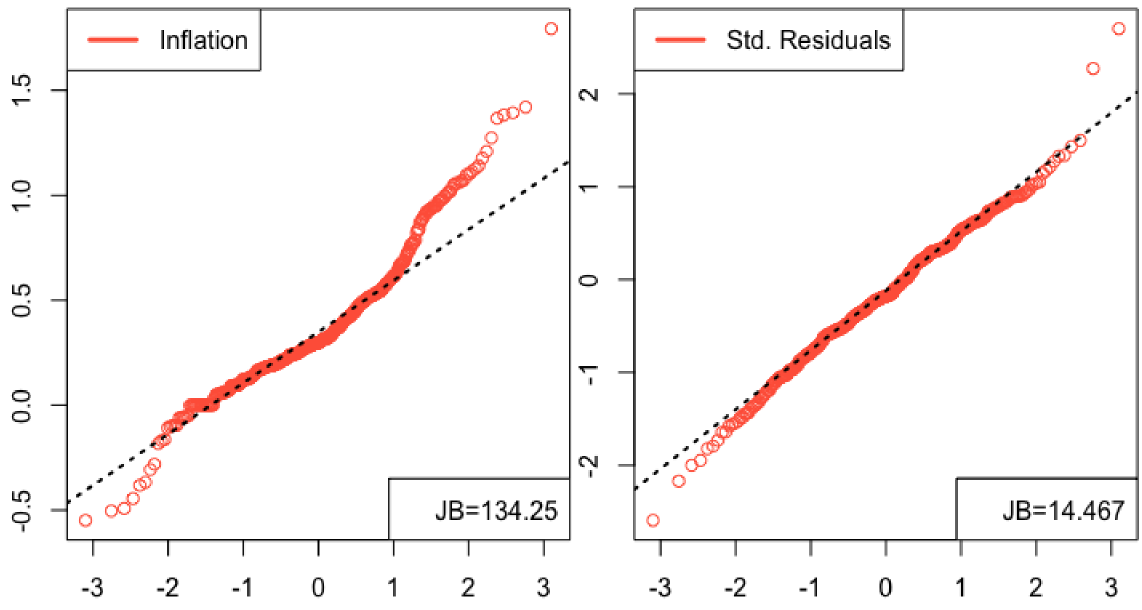

Figure 19: ARMA. Out-of-sample 1-step-ahead forecast

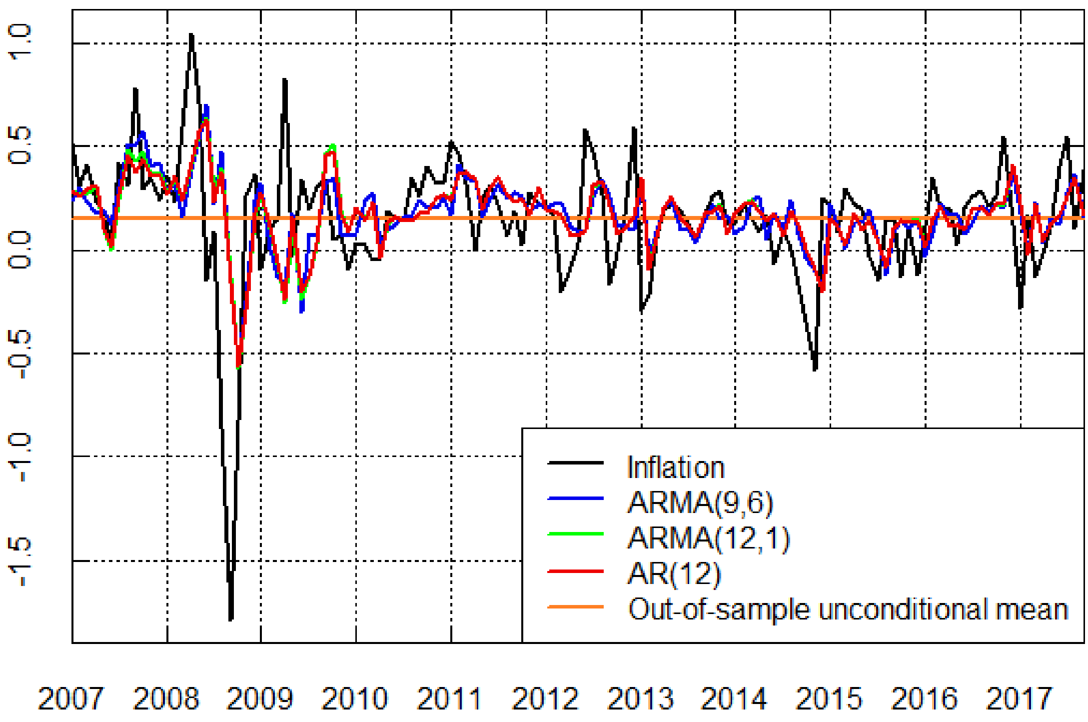

Figure 20: ARMA. Out-of-sample dynamic forecast

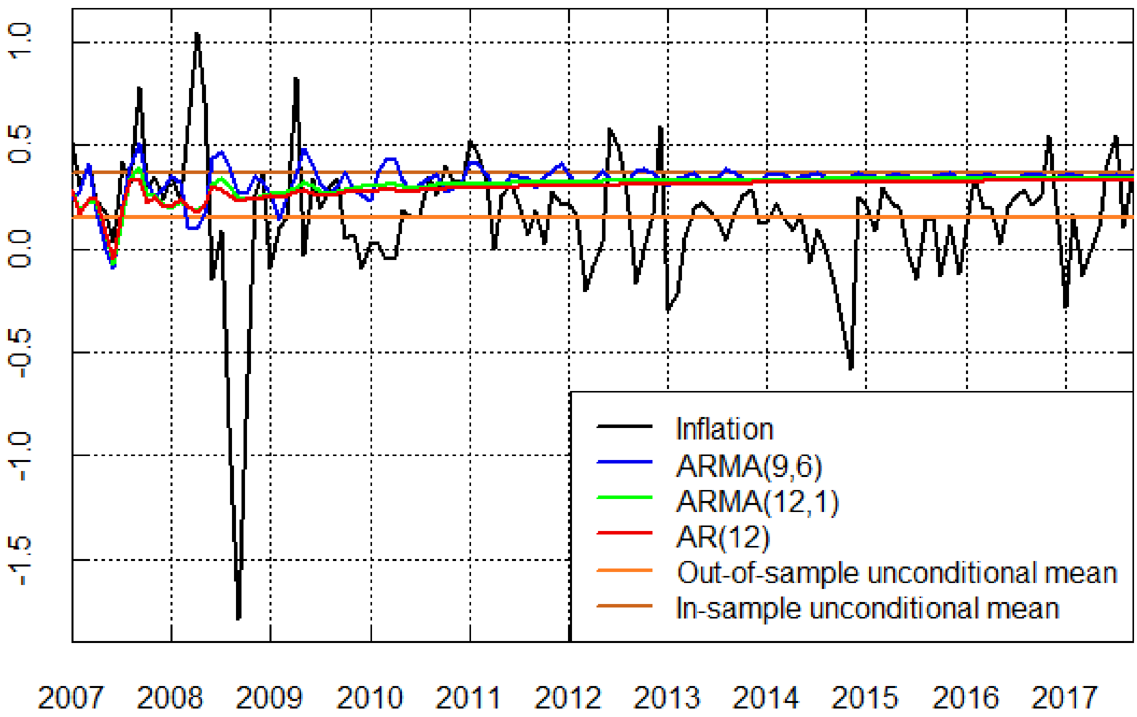

Figure 21: ARCH(3) & GARCH(1,1). Out-of-sample 1-step-ahead forecast