Examining the Dollar-to-Euro Exchange Rate with Time Series Models

Abhay Duggirala, Christina Sorochinsky, Eva Harris, Callan Fang

May 7, 2025

Introduction & Motivation

The Foreign Exchange Market (FX) is a global decentralized market where currencies are traded. In this market, the relative supply and demand of each currency determines its respective value with respect to other currencies. The FX market is by far the largest market in the world by trading volume. Main participants in the market include large international banks (JP Morgan is the top currency trader with 11% market share).

After "liberation day," one of the most important new trends was the great sell-off of the dollar (as well as the spike in U.S. Treasury yields). Given this heightened volatility in the exchange rate for the USD, we wanted to approach this project from the perspective of an investor who seeks arbitrage opportunities. Though we are familiar with the difficulties of forecasting returns and "beating the market" by getting a free lunch (given the Efficient Markets Hypothesis), we sought to fit models that would give us an idea of future market patterns that could be helpful for an investor on the FX market. Moreover, beyond seeking profit, forecasting volatility and (attempting to) forecast returns in this market is quite useful as it can help us understand relative trends in demand and supply of a given currencies, which proxies for trade and capital markets dynamics to some degree. Understanding these trends are both important at a national trade policy level and also important for understanding how certain financial institutions may be affected as the dynamics of the market change.

One of the more interesting markets right now is the USD/Euro FX market. This is because of the massive size of both markets, their high trading volume, and the huge changes each economy is seeing right now (Germany with Chanceller Merz and his "fiscal bazooka" that is likely to have huge effects on the Eurozone and the U.S. with Trump's isolationist trade policies). As such, we focus on the USD/Euro FX market in this project (though the same methods can be applied to any other currency exchange).

Data

The data we pulled for this project come from the Federal Reserve Economic Data's (FRED) U.S. Dollars to Euro Spot Exchange Rate series (DEXUSEU). This data provides us with end of day exchange rates (number of dollars per Euro) every day for the past few decades. Weekends are excluded as the markets are closed on the weekends.

We decided to make a few changes to the series to prepare for our analysis. First, since we are interested in longitudinal changes and trends in the exchange rate rather than daily changes, we decided to transform our data into a weekly average of the exchange rate (Monday-Friday). For us, understanding the general week-over-week changes in the exchange rate is more in tune with our goal to understand general trends in the market (that reflect general global trends) rather than worry about market noise.

When choosing the time frame of our analysis, we decided to start after the end of the Global Financial Crisis (GFC) as there have been significant changes in the market since then and also because the GFC was a highly unusual event that we believe would make our model less generalizable to today's market. As such, our dataset spans from July 1st, 2009 to May 2nd, 2025.

Lastly, we log-differenced our time series in order to look at weekly percent change in this exchange rate (in other words, the return to holding Euros every week) as this is the most interpretable series. Our resulting log-differenced dataset had 825 observations.

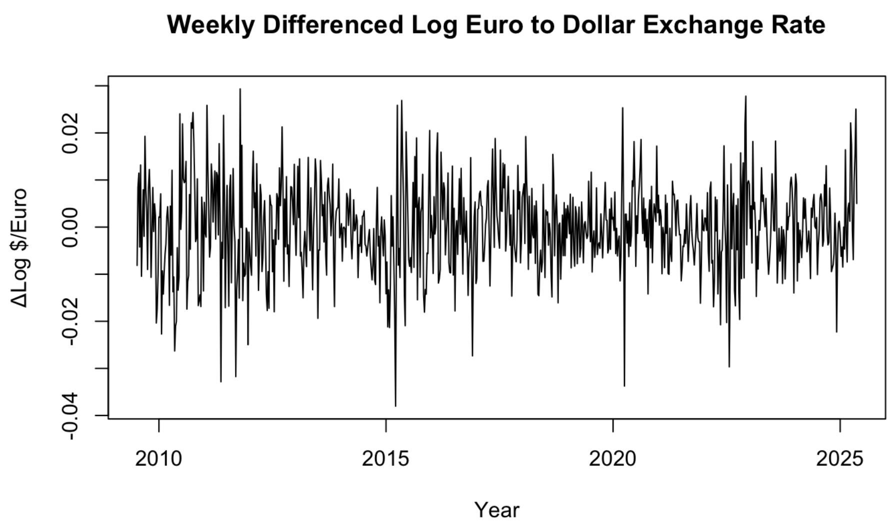


Figure 1: Trace Plot of Our Time Series

Model Selection

Our team went about model selection in the following manner: We had each member of the team try certain models that could plausibly fit to the data, compared and contrasted the benefits, drawbacks, and predictive power of each of these models, and then chose a model that we believed to be the best fit to the data. The model that ultimately was chosen was the ARMA+GARCH model. In the next subsection, we walk through the thought-process behind choosing this model. In the following subsections, we include short descriptions of the other models we tried.

ARMA+GARCH Selection Process

Initially, we believed the GARCH model would be a great place to start. This is because this is the classic model used for forecasting in financial markets for a few reasons: 1) It captures volatility clustering which is a well-documented facet of financial markets through its time-varying conditional variance, 2) It follows the Efficient Markets Hypothesis in that it does not say anything about returns (without an ARMA part added to it), and 3) by understanding volatility forecasts, risk managers can understand how big the next period's moves may be in order to capture important statistics such as VaR which is important for many hedge funds and traders alike.

If we also take a look at Figure 1, we see that there does appear to be some volatility clustering in our data and it appears to be hard to predict the returns for any given week. As such, GARCH was a must-fit model in our analysis. With this model in mind, we started with the classic pre-model selection actions. This meant that we started by looking at the ACF and PACF of our returns as seen in Figure 2.

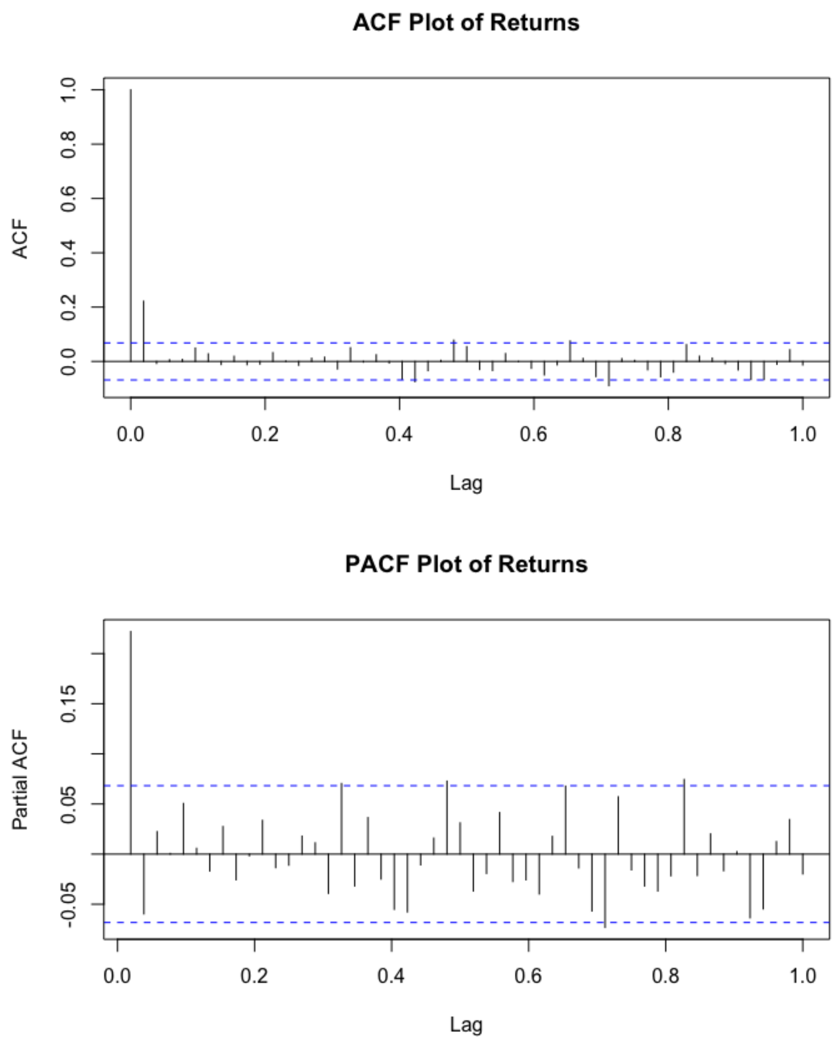


Figure 2: ACF and PACF Plot of Returns

Now, this ACF/PACF plot pair is interesting because it does suggest a significant autocorrelation for 1 week lags (i.e. there is a significant correlation between last week's return and this week's return). This was unexpected, but appears to be pretty evident given the data. We will also see later that fitting an ARMA+GARCH does perform better than a normal GARCH in terms of AIC while still passing the necessary diagnostics. The intuitive analysis is that although returns are often approximately white noise, they can sometimes exhibit short-term serial dependence due to market dynamics or investor behavior.

Given that our ACF cuts off at the first lag (and PACF at lag 1 equals the ACF), we expect that an MA(1) model would be a good initial fit for the data. For an ARMA+GARCH model fit, we first look to see which ARMA model would be the best fit to the data and then from there investigate the ACF/PACF of the squared/absolute residuals. As such, we fit an ARMA(0,1), ARMA(1,0), and ARMA(1,1), and see which model performs the best on our time series. As expected, the ARMA(0,1) or MA(1) performed the best in terms of AIC (-5395 compared to -5392 compared to -5393). Figure 3 illustrates the diagnostic plots for this fit and Figure A.1 illustrates the diagnostics for the AR(1) and ARMA(1,1) models which are similar (with the exception of the AR(1) Ljung-Box plot having values that are much closer to being significant).

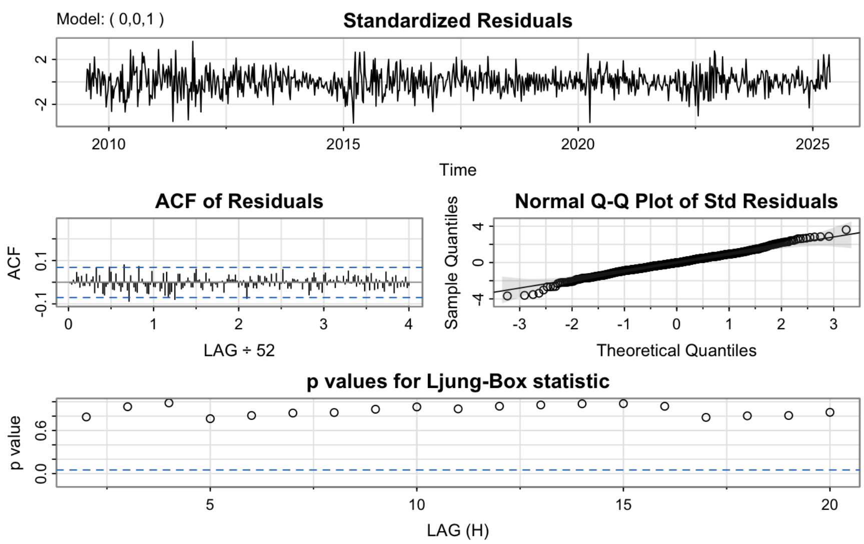
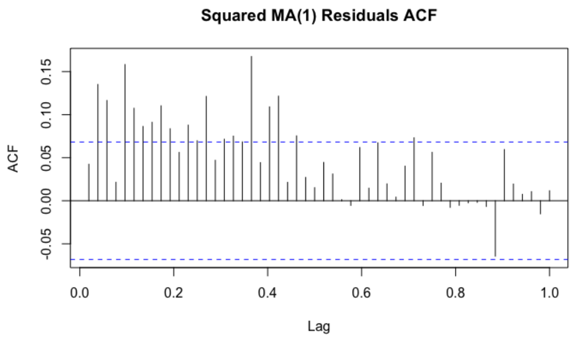


Figure 3: MA(1) Diagnostic Plots

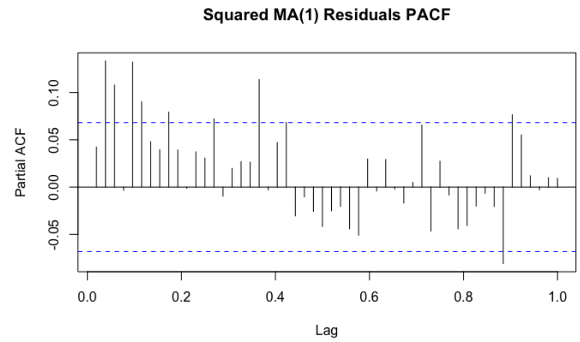
These diagnostic plots for MA(1) illustrate no more correlation in our residuals (ACF and Ljung-Box plot) and normality appears reasonable meaning we can trust the inference for this model. Our coefficients for this ARMA(0,1) are $\theta_1 = 0.2396$ with a p-value of $p = 0$ and an intercept that is insignificant. As such, we initially have (before fitting our GARCH part to estimate how our W_t are constructed based on our conditional variance):

$$r_t = W_t + 0.2396W_{t-1}$$

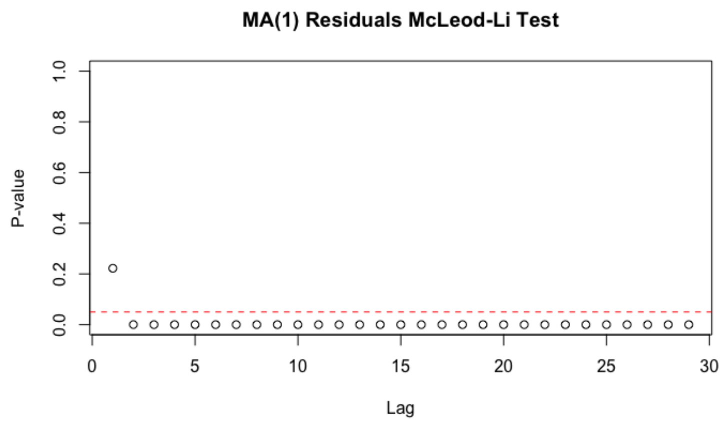
Now, it is time to check for ARCH/GARCH effects by investigating the ACF and PACF of the squared residuals and by also using the McLeod-Li Test (a way to combine ACFs of different lags to help us see autocorrelation of squared residuals) as seen in the below Figure.



(a) MA(1) Squared Residuals ACF



(b) MA(1) Squared Residuals PACF



(c) McLeod-Li Test

These figures illustrate some significant GARCH effects. We see significant correlation between our squared residuals that does not taper off quickly as evidenced by the significant PACF, ACF, and McLeod-Li values (which would suggest more of an ARCH effect). As such, it appears that it would be prudent to fit a GARCH model for the variance of our returns time series. The main question that now comes up is what type of GARCH model would be best combined with the MA(1) model. GARCH(1,1) is typically the gold standard for this sort of financial forecasting as it is most parsimonious GARCH model. We also will try the GARCH(1,2), GARCH(2,1), and GARCH(2,2) models. Figure A.6 provides us our EACF for our squared residuals, but it is not super informative as we do not get the triangle of zeroes that we typically hope for until very high orders of GARCH are reached. However, given our prior knowledge, we are not going to overinterpret the ACF here and will not choose a high order suggestion and rather stick with the parsimonious models rather than fit to the noise.

AICs for ARMA+GARCH Models:

- ARMA(0,1) + GARCH(1,1): -5464
- ARMA(0,1) + GARCH(1,2): -5452
- ARMA(0,1) + GARCH(2,1): -5471
- ARMA(0,1) + GARCH(2,2): -5468

Using AIC to choose the best ARMA+GARCH model, we see that ARMA(0,1) + GARCH(2,1) is the best model. It is also worth noting that when running the McLeod-Li Test on our ARMA(0,1) + GARCH(1,1) (our standard expected model), we have significant values as shown in Figure A.3 meaning this more parsimonious model would have failed our necessary diagnostics (there is remaining conditional heteroskedasticity/ARCH effect in the residuals). It is also simply a worse model. As such, we decide to go with the ARMA(0,1) + GARCH(2,1) as our final model for this analysis.

Other Models Tried

GARCH and IGARCH

One of the models we wanted to test was the IGARCH model. Given that we had significant lags over very long periods (as we saw in our MA(1) residuals ACF/PACF and our McLeod-Li test), we hypothesized that there could be an angle to use the IGARCH model given some potentially very long range dependence in the volatility/variance of our returns. If we also look at our ARMA(0,1) + GARCH(2,1) coefficients (in the next section), we see that our alpha and beta coefficients nearly sum to 1 (0.98). As such, using this logic, we fit an ARMA(0,1) + IGARCH(1,1), ARMA(0,1) + IGARCH(2,1), ARMA(0,1) + IGARCH(1,2), and ARMA(0,1) + IGARCH(2,2) on our returns data.

We found a few main results. 1) None of these models had a higher AIC than our best ARMA(0,1) + GARCH(2,1) model. 2) All models except for the ARMA(0,1) + IGARCH(1,1) had all coefficients other than the MA(1) coefficient insignificant (under a student-t distribution as our residuals are not normal as we see in the next section). See Figure A.4 for our estimated coefficients for this model. 3) When we look at the diagnostics of the ARMO(0,1) + IGARCH(1,1) model (the only model with significant coefficients), we see significant values for the McLeod Li Test at lower lags which suggests there are continued remaining ARCH effects and the model is underspecified as seen in Figure A.5. However, more complex IGARCH models exhibit no statistical significance, so these models are also not good to use. As such, given the worse fit, and issues with diagnostics and significance, it is evident that the ARMA + IGARCH model is not the best for the data.

Intuitively, this makes sense. We do not really expect very long-range dependence of volatility of markets. Volatility usually fades and volatility many weeks ago should not be associated with volatility that is upcoming. As such, an IGARCH model for the variance of our returns is not something that we would intuitively want to use given our knowledge of the FX market.

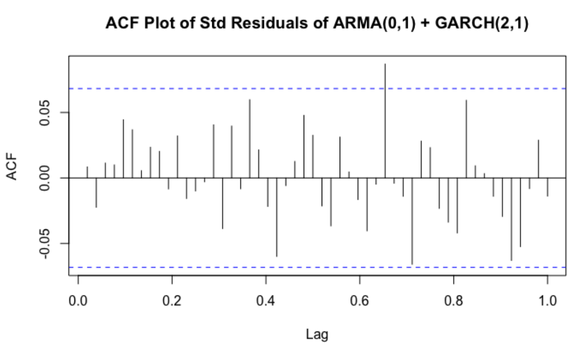
It is also worth noting that we tried to fit a GARCH (by itself) on the returns data. We fit GARCH(1,1), GARCH(2,1), GARCH(1,2), and GARCH(2,2) based on the EACF of our returns data squared in Figure A.6. Our returns data showed significant squared return dependence (McLeod-Li and PACF/ACF of squared returns that looked quite similar to the MA(1) residuals squared from earlier). We found that the GARCH(1,1) modeled had the lowest AIC of -5418 which was still weaker than the ARMA + GARCH models. In addition, although the Generalized Portmanteau Test and McLeod-Li Test both showed showed insignificant values (means there are no more ARCH effects in the data), inspecting the ACF of our residuals shows that it does not mirror white noise given a significant value at lag 1 (which we removed with our MA(1) series in the ARMA+GARCH model). All of this is illustrated in Figure A.7. As such, we decided that the ARMA+GARCH model is still the best model in this case.

Lastly, a random walk with drift model was explored as a possible specification. However, we saw that our model with only an intercept term was not statistically significant given a p-value of 0.442 which is to be expected as the differencing + logging of our data would in effect remove the general trend. In addition, the Ljung-Box plot for both regular and squared residuals showed significant correlation (meaning both the MA(1)-based correlation and conditional variance ARCH effects were not accounted for in this model). The model also had an AIC of -5352.93.

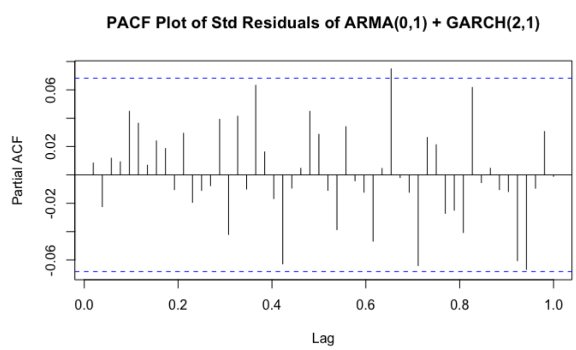
In addition, during the exploratory phase of this project, we tried fitting ARIMA models to this time series. The best ARIMA model (ARIMA(1,1,2)) had an AIC of -5370 and did not account for the significant correlation of squared residuals. In other words, these ARIMA models did not account for the volatility clustering we see in our data and did not fit the data as well. Going through the process of fitting the ARIMA still did provide us with the finding that first order differencing of our log series was best for this time series (second order differencing showed signs of overdifferencing given a significantly negative first lag in the ACF plot).

ARMA(0,1) + GARCH(2,1) Model Diagnostics

Now, we move onto model diagnostics for our chosen model. The goal is to look at the standardized residuals of our model and investigate whether they appear IID. We do this by using the ACF/PACF of these standardized residuals, running the generalized portmanteau test, looking at the McLeod-Li test (on our squared standardized residuals to see if we still see ARCH effects), and also checking normality to ensure our inference is correct. Below, we have our ACF and PACF of our standardized residuals:



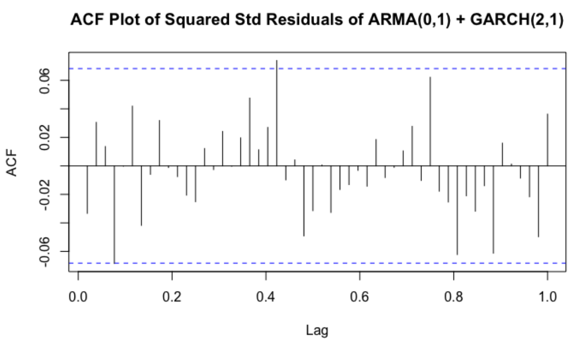
(a) PACF of Chosen Model Std Residuals



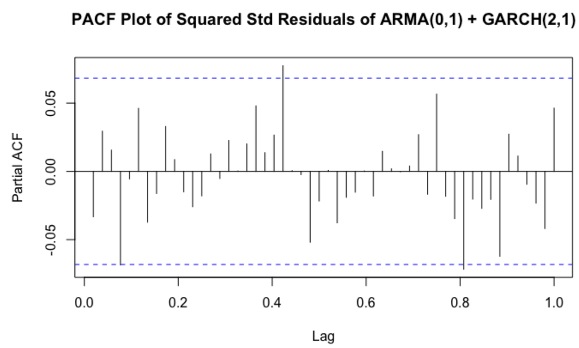
(b) ACF of Chosen Model Std Residuals

We see that these plots show no significant autocorrelation, which is necessarily in an IID series. As such, our model passes this test. Now, we check if we see that the standardized residuals still look like IID noise in the second moment by looking at the gBox p-value, the McLeod-Li test, and ACF/PACF of our squared standardized residuals. Our Box test gives us a p-value of 0.8579 (using squared method) which tells us there appears to be no significant correlation in our squared residuals. We also see the PACF, ACF, and McLeod-Li Test below all suggest the same conclusion.

Now, the last assumption to check is normality. This is only really required to ensure that our inference is statistically sound. Otherwise, the model is well-specified based on these diagnostics. We find from the model's outputted Jarque-Bera test that $p = 0.06$ which is quite close to significant at the 5% level (reject the null hypothesis that the standardized residuals come from a normal distribution). The normal QQ plot of our standardized residuals in Figure 7 also shows quite a heavy lower tail which again seems to suggest that the normal distribution is insufficient to describe the behavior of our standardized residuals. Because of this, we decide to refit the model using a student-t distribution to account for these heavy tails. This is the model that is outputted in the next section. With diagnostics passed, we move onto the final model and its forecasts.



(a) ACF of Squared Std Residuals



(b) PACF of Squared Std Residuals

(c) McLeod-Li Test

Estimated Model and Forecasts

Estimated Model and Interpretation

We rerun the fit with a student-t distribution as the density of ϵ_t. The fitted model with standard-errors in parentheses (assuming a mean of 0 given our mean parameter is 0.00035 with a p-value of 0.31) is given below:

$$X_t = W_t + 0.25_{0.033} W_{t-1}$$

where

$$W_t = \sigma_t \epsilon_t \quad e_t \sim_{iid} t_{10}$$

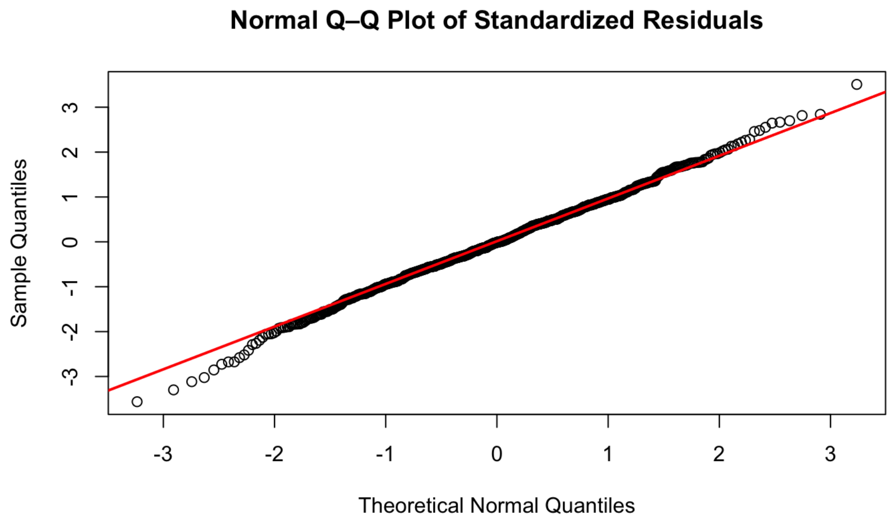


Figure 7: Normal QQ Plot of Standardized Residuals

and

$$\sigma_t^2 = 0.0000019_{9.658e-07} + 0 \cdot W_{t-1}^2 + 0.092_{0.0495} W_{t-2}^2 + 0.89_{0.026} \sigma_{t-1}^2$$

Our output from R is in Figure A.8 for reference. The coefficients that are significant in this model (at a 10% level) are the θ_1, b_2 (our second lag ARCH effect), and a_1 (our first lag GARCH effect). b_0, our constant term in our conditional variance equation is also significant. The shape parameter for the t-distribution of ϵ_t is 10 and is also significant. The first ARCH effect b_1 is insignificant and so is the μ (intercept) term as mentioned earlier.

The interpretation of this model is as follows. The MA(1) coefficient tells us that successive weekly moves tend to have the same sign slightly more often than chance. This is interesting given the difficulty in forecasting returns, but it probably has to do with large institutional investor behavior, geopolitical factors, or other predictable investor behavior. The insignificant mean tells us that the average weekly return is indistinguishable from zero. This is consistent with weak-form market efficiency (no free α). Interestingly, we find that the most recent shock (ARCH lag 1 coefficient) has no incremental impact on next week's volatility and that the shock two weeks ago still affects today's volatility. This is fascinating as it is pretty unexpected and we do not have an intuitive explanation for why this may happen. We could hypothesize that this is because information accumulation occurs over a week long period or because there is some additional noise in the data the model fits to. The GARCH coefficient being so large at 0.891 tells us that volatility is very persistent and lasts many weeks ($b_1 + b_2 + a_1 = 0.98$). Shocks decay slowly and volatility clusters for a while which is quite interesting (and motivated us trying the IGARCH model earlier). We can calculate our unconditional variance

$$\hat{\sigma}^2 = \frac{b_0}{1 - b_1 - b_2 - a_1} = 1.1 \times 10^{-4}$$

showing how our long-run unconditional weekly standard deviation sits around 1%. Lastly, our student-t shape parameter tells us that extreme moves are far more likely than under the normal assumption which makes sense in financial markets where tail events are fairly common.

The long-run persistence of volatility tells us that volatility regimes dominate forecasts for a significant number of weeks which tells us that as investors, it is important to monitor "volatility regime shifts" over time. This is pretty surprising given our prior knowledge that suggests that volatility does not last this long in the stock market, but it may be that the idea of "volatility regimes" is prevalent in the FX market especially given its greater dependence on geopolitics and country trends. Moreover, even though slight positive autocorrelation may offer marginal timing edges, it is likely that transaction costs in FX swamp the signal, removing our opportunity for a "free lunch." However, the volatility forecasting is still valuable for risk-management purposes.

Forecasts

With our fitted model, we first look at how the model fit on our full time series. Figure 8 illustrates our model fit on the data. It appears that conditional variance estimates do rise during periods of volatility, showing that the desired volatility clustering shows up in the data. In addition, the conditional mean estimates seem to generally center around zero and move more during periods of high volatility as expected given our MA(1) term. All in all, the model appears to approximate the general volatility and return trends USD/Euro FX market pretty well given the constraints and difficulty in forecasting this market.

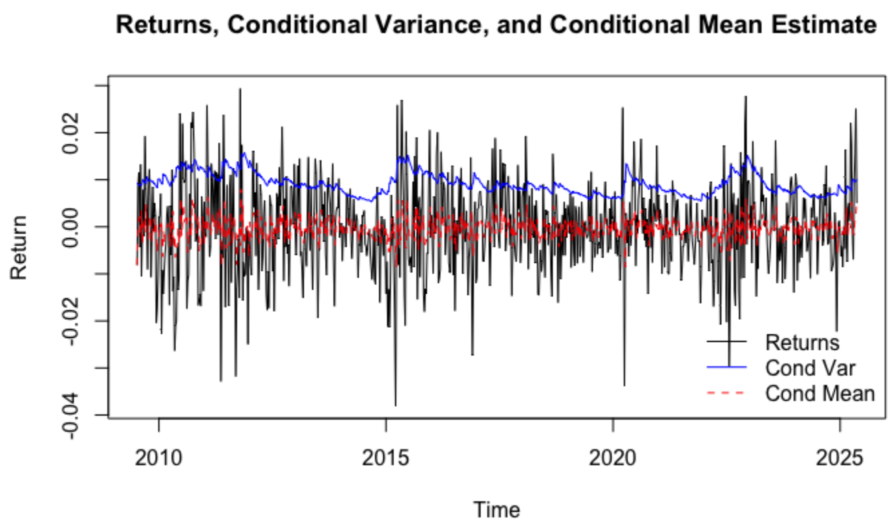


Figure 8: Trace Plot with Fitted Values for ARMA(0,1) + GARCH(2,1)

Now, we turn to forecasts. With an MA(1) model, we know that the mean forecast for one step ahead is $\hat{r}_{T+1} = -\sum_{j=1}^{\infty}(-\theta_1)^j r_{T+1-j}$. For more than one step ahead, for an MA(1) model, the prediction is zero (and with our nonzero mean, it is μ). Our forecast of conditional variance is detailed in Lecture 11 and we use the R command predict to receive these values. Figure 9 illustrates the next 10 weeks of forecasted returns and their standard deviations for our model along with the last 20 observations (see Table A.1 for the exact forecast values).

As we see, the mean forecast is slightly different for the first forecasted value and then returns to the mean of the fitted model. We see that standard deviation remains elevated for a few weeks

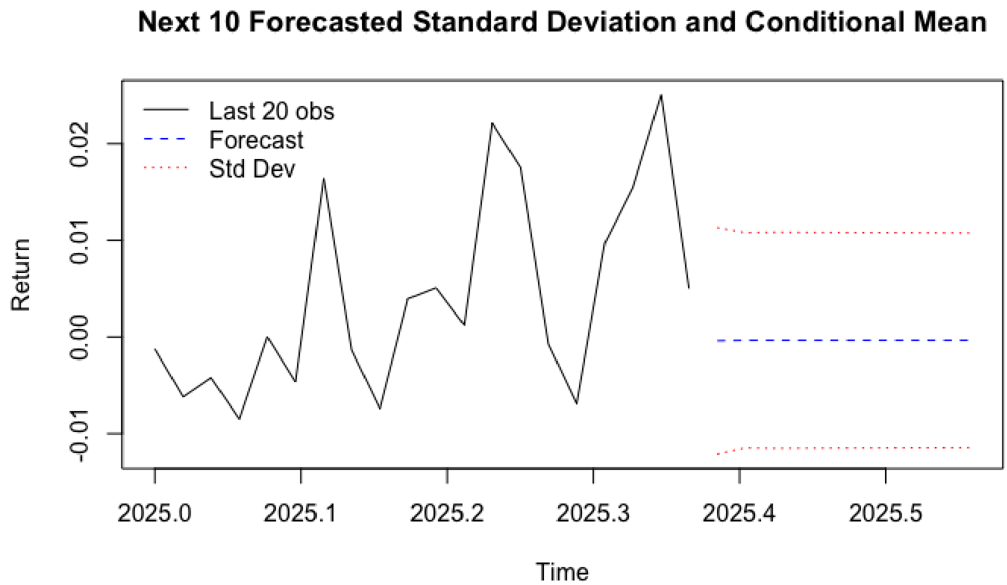


Figure 9: Last 20 Observations Superimposed on Next 10 Forecasted Values

(volatility persistence telling us there is increased volatility in the market currently) and begins moving toward the long-run unconditional standard deviation of $\sim 1\%$ as the forecast horizon gets longer. Volatility does appear to persist over a significant period of time (as further evidenced in Table A.1). However, it appears that as the forecast gets longer in horizon, it becomes less accurate. In general, this forecast model can help us understand week-to-week movements in the USD/Euro FX market while helping us understand volatility over a short forecast horizon of a few weeks.

Discussions and Conclusions

Summary of Findings

Using 825 weekly log-returns on the USD/EUR exchange rate (2009-07-01 to 2025-05-02), we compared various time series specifications and chose an ARMA(0,1) + GARCH(2,1) with student-t error distribution as the preferred model.

- The MA(1) term captured some week to week serial correlation we found in the ACF.
- The second ARCH lag plus a very persistent GARCH term explained persistent volatility clustering.
- Normal error diagnostics were marginal and thus switching to a t distribution worked to eliminate issues with residuals while retaining our strong diagnostics.
- Other models such as the ARMA + IGARCH model did not perform as well as this model.

Together, these results indicate that weekly FX volatility is persistent, while mean returns are hard to distinguish from zero which consistent with weak-form market efficiency.

Strengths of the Analysis and Opportunities for Improvement

Our analysis was relatively strong, but could have been improved if given more time. Our model search was fairly comprehensive as we compared our chosen model with IGARCH, higher order GARCH, and random walk alternatives before choosing our final model. We also performed rigorous diagnostics which reduced any risk of underspecification. Further, we provided economic interpretation for our parameter estimates which is important in the context of trading (understanding overall macro trends from our models is better than simply just plugging in models).

There were many opportunities for improvement, however. One, we did not have any out-of-sample evaluation. Even though our model selection criterion (AIC) limits overfitting, we did not specifically quantify predictive value through out-of-sample evaluation. We could have performed rolling-window or expanding-window forecasts and compare those to the naive forecast to better understand the performance of the value. In addition, we could have better explained variance by incorporating other exogenous macro variables. Interest-rate spreads, Fed/ECB rate announcements, VIX, and so on into an ARMAX-GARCH or VAR-GARCH framework to build a more predictive model. It also would have been interesting to try a daily forecasting model to see how the model may differ for high frequency trading. Furthermore, given our idea of the "volatility regime," a Markov-Switching model may also better fit our data as these may handle regime shifts like the euro-area debt crisis of the pandemic period. Overall, this model is a good starting point, but much can be done to improve this model to compete with what many large quantitative hedge funds use today.

Appendix

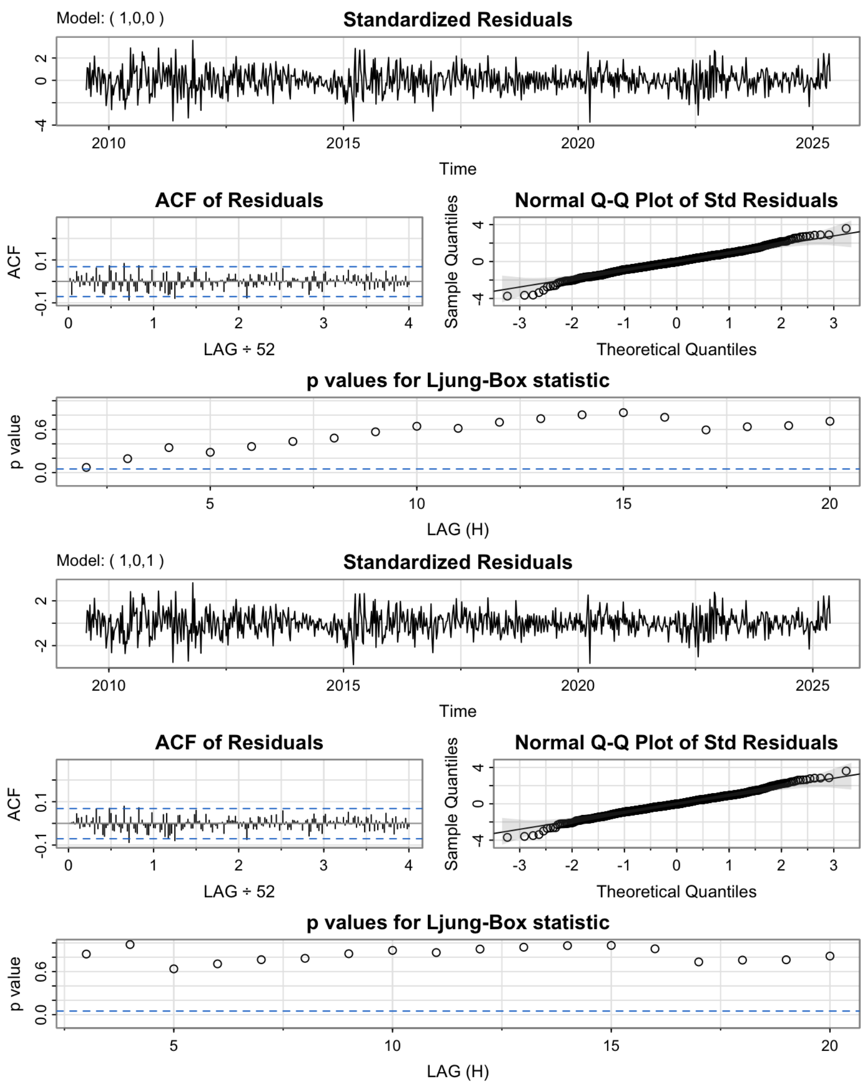


Figure A.1: AR(1) and ARMA(1,1) Diagnostic Plots

AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	x	o	x	x	x	x	x	x	o	x	o	x
1	x	o	x	o	x	o	o	o	o	o	o	o	o	x
2	x	x	o	x	x	o	o	o	o	o	o	o	o	o
3	o	x	x	o	x	o	o	o	o	o	o	o	o	o
4	o	x	x	x	o	o	o	o	o	o	o	o	o	o
5	x	o	x	x	x	o	o	o	o	o	o	o	o	o
6	x	x	o	x	x	x	o	o	o	o	o	o	o	o
7	x	x	x	x	x	x	o	o	o	o	o	o	o	o

Figure A.2: EACF for Squared Residuals from MA(1) Model

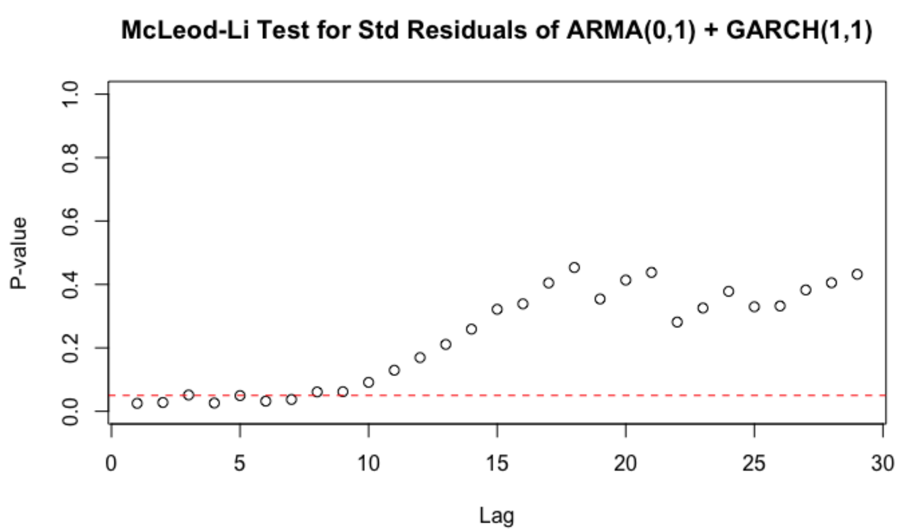


Figure A.3: ARMA(0,1) + GARCH(1,1) McLeod-Li Test

	Estimate	Std. Error	t value	Pr(>|t|)
mu	-0.000331	0.000348	-0.95108	0.341562
ma1	0.243278	0.036158	6.72827	0.000000
omega	0.000001	0.000002	0.40082	0.688551
alpha1	0.082208	0.024371	3.37322	0.000743
beta1	0.917792	NA	NA	NA
shape	12.239727	4.702274	2.60294	0.009243

Figure A.4: ARMA(0,1) + IGARCH(1,1) Coefficients (Student-t Distribution)

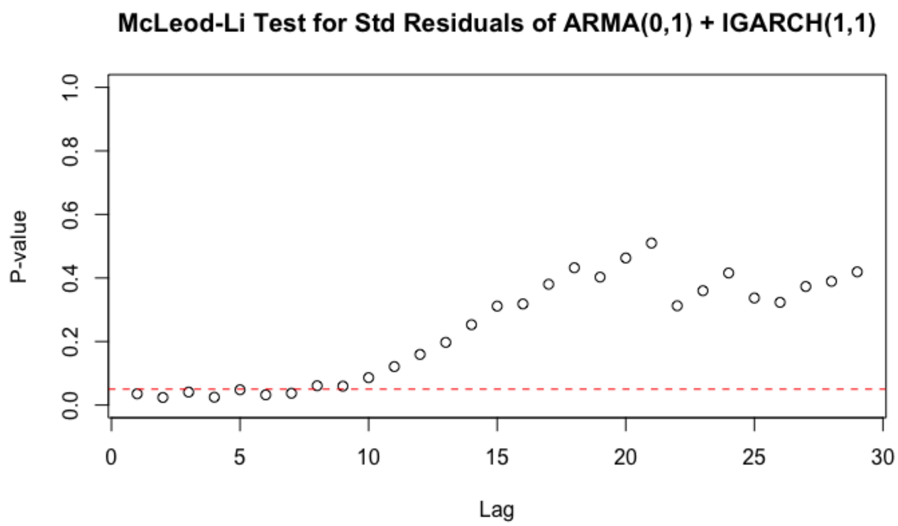


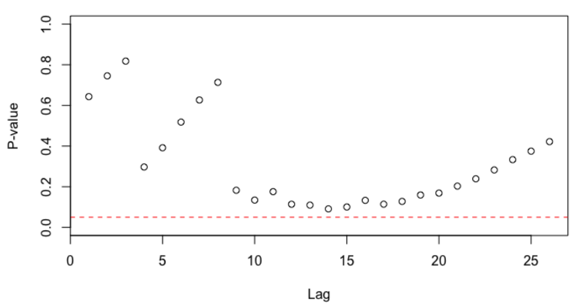
Figure A.5: ARMA(0,1) + IGARCH(1,1) Mc-Leod-Li Test

Mean Forecast	**Std. Deviation**
-0.0003960333	0.01172055
-0.0003507571	0.01114408
-0.0003507571	0.01118727
-0.0003507571	0.01117117
-0.0003507571	0.01116081
-0.0003507571	0.01115008
-0.0003507571	0.01113955
-0.0003507571	0.01112917
-0.0003507571	0.01111895
-0.0003507571	0.01110887

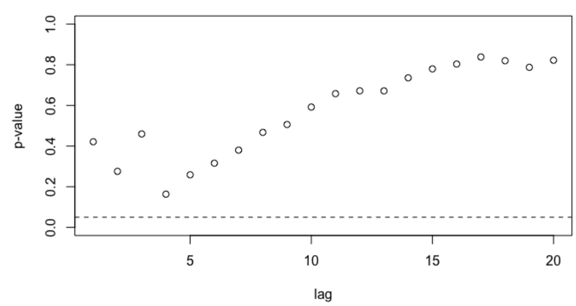
Table A.1: Mean Forecasts and Standard Deviations from the ARMA(0,1) + GARCH(2,1) Fit

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	o	x	x	x	x	x	o	x	x	o	x
1	x	o	o	o	x	o	o	o	o	o	o	o	o	o
2	x	x	o	o	o	o	o	o	o	o	o	o	o	o
3	o	o	x	x	x	o	o	o	o	o	o	o	o	o
4	o	x	x	x	o	o	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o	o
6	x	x	o	o	x	x	x	o	o	o	o	o	o	o
7	x	x	o	o	x	o	x	o	o	o	o	o	o	o

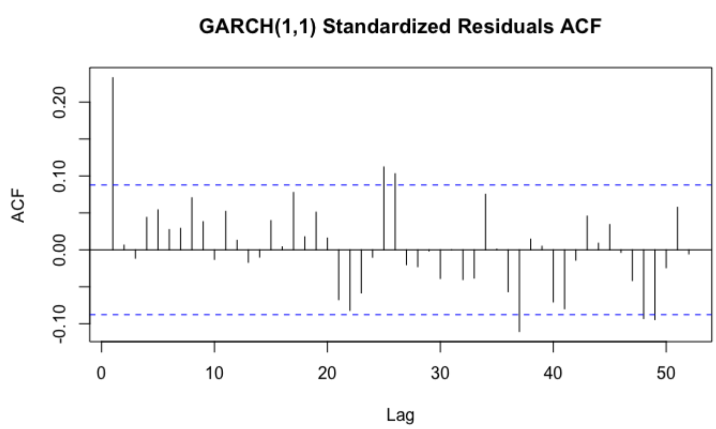
Figure A.6: EACF of Squared Returns Data



(a) GARCH(1,1) McLeod-Li Box Test



(b) Generalized Portmanteau Test (Squared)



(c) GARCH(1,1) Residuals ACF

```
        Estimate  Std. Error  t value Pr(>|t|)
mu     -3.508e-04   3.487e-04   -1.006 0.314409
ma1     2.498e-01   3.339e-02    7.483 7.28e-14 ***
omega   1.851e-06   9.658e-07    1.917 0.055233 .
alpha1  1.000e-08   4.318e-02    0.000 1.000000
alpha2  9.249e-02   4.954e-02    1.867 0.061921 .
beta1   8.906e-01   2.561e-02   34.780  < 2e-16 ***
shape   1.000e+01   2.597e+00    3.851 0.000118 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.8: Coefficient Output for ARMA(0,1) + GARCH(2,1)