

ECON3350: Applied Econometrics for Macroeconomics and Finance

Research Report 1

Question 1

(a) Plots and description

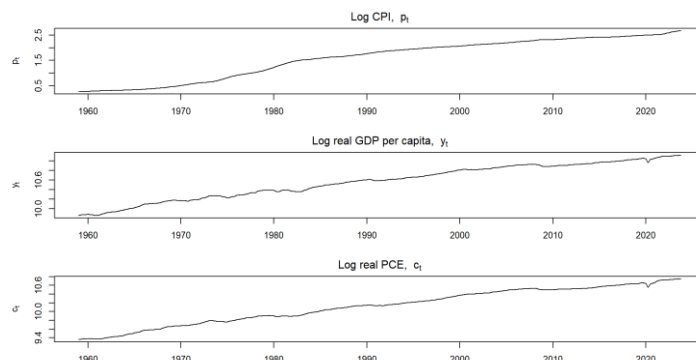


Figure 1. Log levels of the US Consumer Price Index (p_t), real GDP per capita (y_t), and real personal consumption expenditure (c_t), 1959Q1 to 2023Q4.

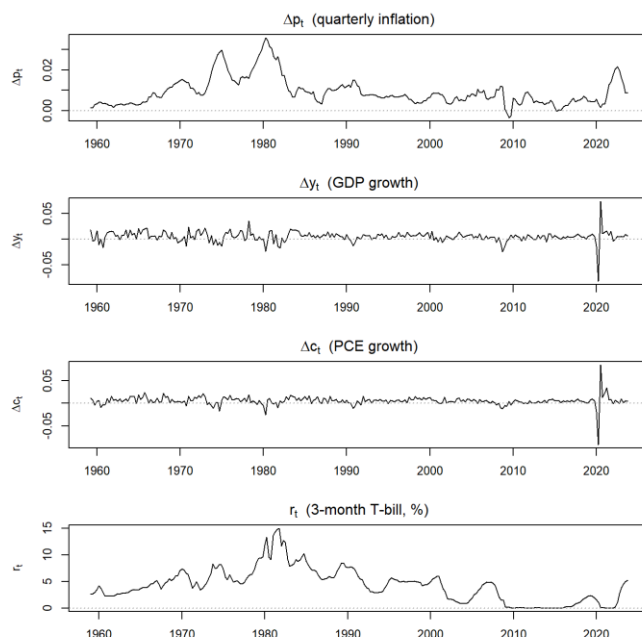


Figure 2. Quarterly log-differences Δp_t , Δy_t , and Δc_t , and the 3-month Treasury-bill rate r_t , 1959Q1 to 2023Q4.

Figure 1 shows strong upward movement in p_t , y_t , and c_t , so the level series do not look covariance-stationary. The rise in p_t is steepest from the late 1960s to the early 1980s, consistent with the high-inflation era. By contrast, y_t and c_t trend upward more smoothly and move closely together, with visible cyclical slowdowns around 2008-09 and a sharp fall-and-rebound in 2020.

In Figure 2, Δp_t , Δy_t , and Δc_t fluctuate around roughly stable means, suggesting that first differencing removes most of the non-stationarity in levels. Inflation is clearly more volatile in the 1970s and early 1980s and again after 2020. Output and consumption growth are comparatively stable except for the extreme COVID shock in 2020Q2-Q3.

The interest rate r_t behaves differently. It is highly persistent, rises into the early 1980s, then declines for decades before sitting near the zero lower bound after the GFC and again in 2020-21. Visually, it looks much less stable than the differenced real-activity series and may contain substantial low-frequency shifts.

(b) Regressions and means

The following regressions and mean estimates were developed using data up to 2023Q4:

i)

Parameter / series	Estimated intercept ($\mu^\wedge$)	Estimated slope ($\delta^\wedge$)	Estimated regression
p_t	0.2706778	0.010124945	$p_t = 0.2707 + 0.0101t + \varepsilon_t$
y_t	9.9448533	0.004757206	$y_t = 9.9449 + 0.0048t + \varepsilon_t$
c_t	9.4290009	0.005305329	$c_t = 9.4290 + 0.0053t + \varepsilon_t$

ii)

c	Estimated mean ($\mu^\wedge$)	Quarterly (%)	Annualised (%)
Δp_t	0.009310287	0.93	3.72
Δy_t	0.004918979	0.49	1.97
Δc_t	0.005378509	0.54	2.15

The estimated slopes are useful descriptive measures of average growth, but their usual OLS standard errors are not reliable if the level series are non-stationary. I therefore interpret the coefficients as summary measures of trend rather than as the basis for formal t-tests.

(iii) Rationale and comparison

Both approaches estimate long-run average growth, but they come from different assumptions about the data-generating process. Regressing the level on a linear time trend treats the series as trend-stationary and uses the slope to measure deterministic growth.

$$x_t = \mu + \delta_t + \varepsilon_t,$$

Estimating the mean of first differences instead treats the series as difference-stationary and uses the average change to measure drift.

$$\Delta x_t = \mu + \eta_t,$$

In macroeconomic data, both views are plausible: levels are often non-stationary, while first differences are much closer to stationary.

The two estimates therefore target the same underlying quantity - average growth. Their closeness is mainly a consistency check, not a formal test. Distinguishing between the two models requires the unit-root evidence developed in Question 2.

For y_t and c_t , the trend coefficients and mean differences are very similar. For p_t , the gap is larger because a single linear trend puts substantial weight on the high-inflation 1970s and early 1980s, whereas the mean of Δp_t reflects the average quarter-to-quarter change over the whole sample. That instability makes p_t the series that most clearly needs formal unit-root and ARIMA analysis.

Question 2

(a) Choice of three ARIMA models

ADF results were used to choose the value of d . For $\log P_t$, the test does not reject a unit root in levels, but it does reject after first differencing. This suggests that inflation, Δp_t , can be treated as stationary, so it is modelled with $d = 0$. For r_t , the evidence points to an $I(1)$ process: the level series generally fails to reject a unit root, while the first difference, Δr_t , rejects it clearly.

I then estimated ARIMA (p, d, q) models for $p, q \in \{0, \dots, 5\}$. From these, I kept the models that performed best under AIC and BIC, and then retained only those with residuals that were adequately white at short Ljung-Box horizons. The three preferred models for each series are reported below, ranked by AIC with BIC used as a tiebreaker.

Model	Constant	AIC	BIC	LB p-values, lags 1 to 5
ARIMA(3,0,3)	Constant	-2686.38	-2657.92	0.87, 0.95, 0.99, 0.82, 0.65
ARIMA(3,0,4)	Constant	-2685.21	-2653.20	0.89, 0.89, 0.88, 0.95, 0.79
ARIMA(5,0,3)	Constant	-2685.17	-2649.60	0.99, 0.99, 1.00, 1.00, 1.00

Model	Constant	AIC	BIC	LB p-values, lags 1 to 5
ARIMA(3,1,5)	Constant	480.24	512.25	0.99, 1.00, 1.00, 0.99, 0.96
ARIMA(5,1,4)	Constant	480.26	515.83	0.99, 0.93, 0.87, 0.94, 0.93
ARIMA(5,1,5)	Constant	481.82	520.95	0.90, 0.99, 0.97, 0.99, 0.98

Some lag-10 Ljung-Box p-values are small, but the FAQ notes that higher-lag versions of the test have low power and should be interpreted cautiously. Given the strong short-lag diagnostics and the close forecast paths, I treat the shortlisted models as adequate.

(b)

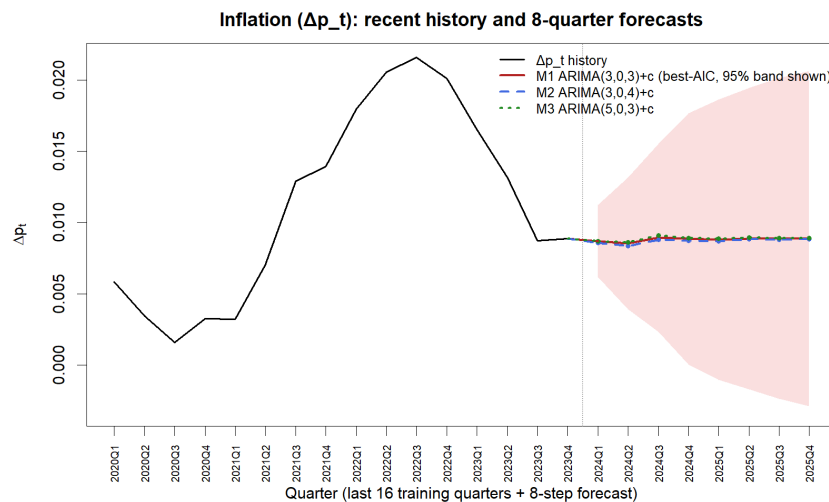


Figure 3. Recent Δp_t and eight-quarter forecasts from the three selected ARIMA models.

(c) Policy uses and uncertainty

These forecasts are useful for policy because they provide a benchmark for where inflation is likely to move. This helps central banks assess how far expected inflation is from target and how strong a policy response may be needed. Here, all three models suggest inflation will settle at around 3.5% annualised, still above the Fed's 2% target.

There is, however, considerable uncertainty around these forecasts. Forecast uncertainty rises with the horizon: for the best-fitting model, the 95% interval is about 1 percentage point annualised one quarter ahead, but widens to almost 5 percentage points by eight quarters ahead. There is also some parameter uncertainty, although model uncertainty is small because the three shortlisted models produce very similar forecast paths.

The main limitation is structural uncertainty: because the model is estimated over a long sample covering very different inflation regimes, its implied long-run mean may not match current conditions well. This is reflected in the tendency of the models to overpredict inflation relative to the realised values, as shown more formally in Question 3 using MSFE, RMSE, MAFE and MFE. ARIMA models also focus on the conditional mean and do not capture changing volatility particularly well, so they are less reliable during periods of instability or unusually large shocks. Overall, these forecasts are best treated as a broad benchmark rather than a stand-alone guide for short-run policy.

Question 3

The 2024Q1 to 2025 Q3 P_t observations were converted into inflation differences and compared with the model forecasts to obtain forecast errors. The three models were then evaluated using MSFE, MAFE, RMSE and MFE, which together summarise forecast accuracy, the typical size of errors, and any overall forecast bias.

Model	Constant	MFE	MAFE	MSFE	RMSE
ARIMA(3,0,3)	Constant	0.001631	0.001631	3.398×10^{-6}	0.001843
ARIMA(3,0,4)	Constant	0.001523	0.001523	3.096×10^{-6}	0.001760
ARIMA(5,0,3)	Constant	0.001696	0.001696	3.618×10^{-6}	0.001902

ARIMA(3,0,4) performs best across all four forecast measures. Its RMSE is 0.00176 per quarter, or about 0.7 percentage points annualised, and its MFE is 0.00152 per quarter, or about 0.6 percentage points annualised. For all three models, MFE and MAFE are the same, showing that inflation was over-predicted in all seven holdout quarters. Since this bias accounts for about 87% of the RMSE, most of the error is systematic rather than random. This suggests that the long-run mean built into the ARIMA models is too high for the 2024-2025 period, reinforcing the structural-break concern discussed in Question 2.

Question 4

(a) Plot of rr_t and stability

Figure 4 shows that rr_t moves widely: around 2-5% in the 1960s, double digits in the Volcker period, near zero after the GFC and again in 2020-21, then rising post-pandemic. Under the assignment's definition, rr_t closely tracks r_t because Δp_t is measured as a quarterly log change and is numerically much smaller. So this proxy is not visibly more stable than the nominal rate, even though it removes some low-frequency inflation movement.

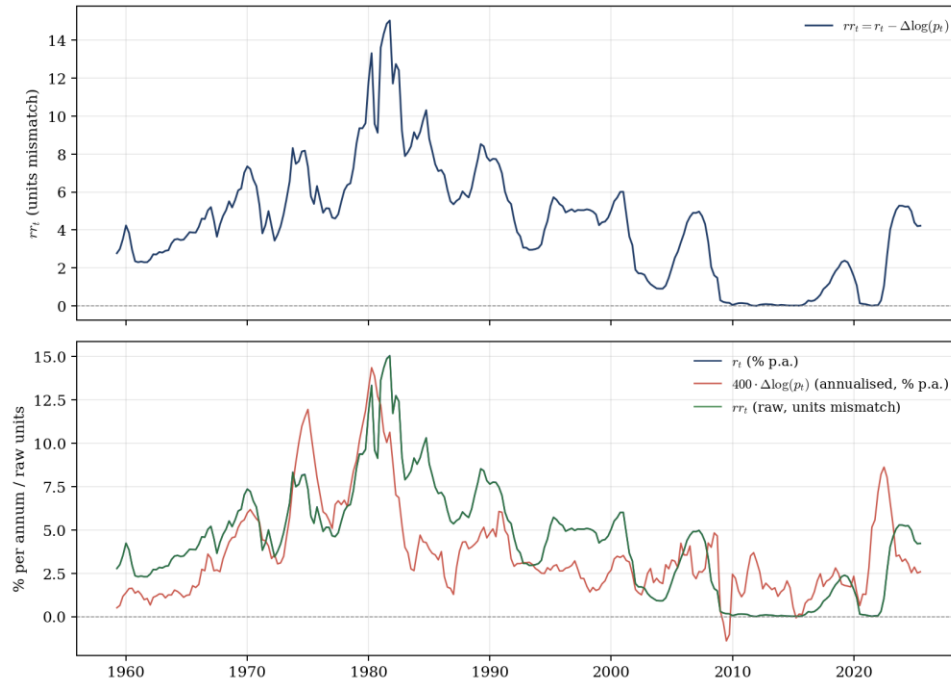


Figure 4. Real-interest-rate proxy $rr_t = r_t - \Delta p_t$ over the full sample, with a comparison to r_t and annualised Δp_t .

(b) Plot of cy_t and dominant feature

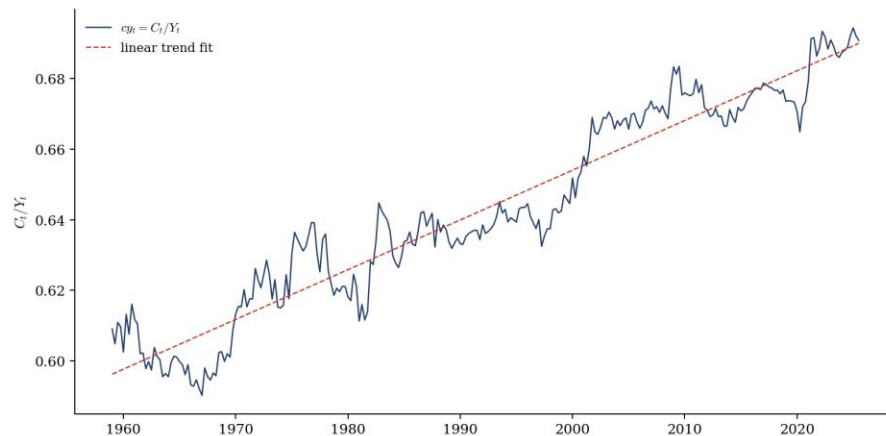


Figure 5. Consumption ratio $cy_t = \frac{C_t}{Y_t}$, 1959Q1 to 2025Q3.

Figure 5 shows a strong upward trend in cy_t , rising from about 0.60 to 0.69 over the sample. The main feature is therefore a persistent increase in consumption's share of output, rather than short-run cyclical noise. A plausible explanation is structural change, including a larger services sector, lower household saving over parts of the sample, population ageing, and a weaker investment share of GDP. Visually, cy_t appears to follow a trend

rather than fluctuate around a fixed mean, although formal testing is needed to determine whether that trend is deterministic or stochastic.

ADF results suggest different treatments for the two series. For rr_t , the evidence against a unit root is weak and disappears at the BIC-selected lag, so I model it conservatively with $d = 1$. For cy_t , the ADF test rejects under the trend specification but not under the versions without a trend, which is the usual pattern for a trend-stationary series. I therefore model cy_t with $d = 0$ and include a linear trend.

(c) ARIMA(7,1,1) for rr_t (no drift)

Parameter	Estimate	SE	t-stat
ar1	-0.361	0.083	-4.36
ar4	0.173	0.064	2.71
ar7	-0.302	0.060	-5.06
ma1	0.863	0.069	12.49

AIC = 479.93, BIC = 512.15, and $\sigma^2 = 0.343$, with the Ljung-Box test failing to reject residual autocorrelation. The choice of $d = 1$ reflects the long and persistent swings visible in Figure 4a. In particular, the deeply negative real rates of the late 1970s and the very high real rates of the early 1980s do not look like temporary fluctuations around a fixed mean. Instead, they suggest that rr_t inherits much of the persistence in r_t . Within this framework, the negative *ar1* term captures some short-run mean reversion, while *ar4* and *ar7* pick up longer cyclical persistence across quarters. The large *ma1* term helps absorb sharp shifts associated with major regime changes. Overall, the model captures the broad movements in the series reasonably well, although it does not account for the clear volatility clustering around 1980. That limitation motivates the later use of a GARCH specification.

(d) ARIMA(1,0,0) for cy_t with constant and linear trend

Parameter	Estimate	SE	t-stat
ar1 (ϕ)	0.876	0.029	29.9
intercept	0.597	0.0036	165.8
trend (qtr)	3.46×10^{-4}	6.4×10^{-5}	5.44

AIC = -2204.43, BIC = -2190.08, and $\sigma^2 = 1.48 \times 10^{-5}$, with the Ljung-Box test again failing to reject residual autocorrelation. The dominant feature of Figure 4b is the steady upward trend in cy_t , rising from about 0.60 in 1960 to about 0.69 in 2025. The estimated trend coefficient of 3.46×10^{-4} per quarter captures this clearly, implying a cumulative increase of roughly 0.09 over the full sample. The high AR(1) coefficient, $\phi = 0.876$, is also consistent with the visual pattern: departures from trend, such as the dip around 2008 and the spike in 2020, fade only gradually rather than disappearing quickly. This supports modelling cy_t with $d = 0$ and a linear trend, implying that the series is trend-stationary: it follows a deterministic upward path, with persistent but ultimately temporary deviations around that trend.

(e) Policy usefulness

Overall, the two models are useful for policy because they capture persistence, mean reversion and short-run adjustment. For rr_t , the model provides a benchmark for how unusual current real interest rates are relative to recent history. This is helpful for judging monetary tightness, debt-servicing pressure, and how sensitive the economy may be to further rate changes. More broadly, the model's forecasts of the conditional mean and variance can also be used to assess uncertainty around policy settings and debt-sustainability risks.

For cy_t , the model is useful because it separates the long-run upward trend from shorter-run fluctuations around that trend. This helps distinguish structural changes in household behaviour from temporary cyclical movements, which is relevant for fiscal and macroprudential policy. Large forecast errors relative to the trend may also signal broader regime changes or shifts in consumption behaviour that matter for policy.

These models are still only approximations. In particular, the higher-order rr_t specification raises some concern about overfitting, and neither model deals well with time-varying volatility. As a result, they are most useful as benchmarks for interpretation and policy discussion, rather than as precise guides during highly volatile periods.

Question 5

(a) Sample variance

Series	σ^2 (sample variance)	Excess kurtosis
CNY	0.337	13.6
USD	0.438	12.2
TWI	0.268	15.0
SDR	0.449	12.5

Daily returns have means close to zero, but they are clearly fat-tailed rather than normally distributed. TWI is the least volatile ($\sigma^2 = 0.268$), which is consistent with its construction as a weighted basket across Australia's trading partners. CNY (0.337) is less volatile than USD (0.438), which fits with the more managed nature of the renminbi. SDR (0.449) is the most volatile, likely because it reflects a basket of major currencies; EUR, GBP, USD and JPY, and therefore captures broader global currency movements. Excess kurtosis of around 12 to 15 reinforces the evidence of fat tails, and the Jarque-Bera test strongly rejects normality. These unconditional variance estimates are useful descriptively, but they do not capture time-varying volatility, which is why a GARCH framework is needed in Question 6.

(b) Plot of $|e_{j,t}|$

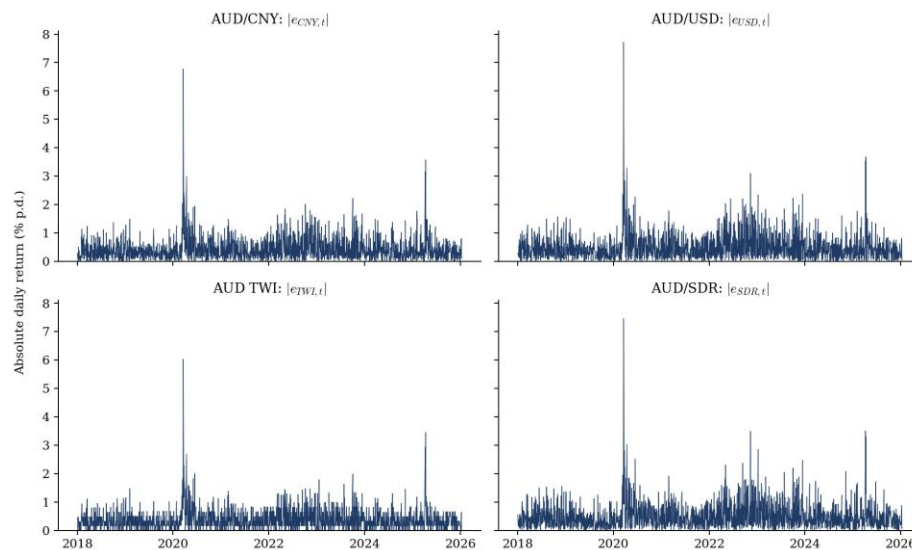


Figure 6. Absolute returns $|e_{j,t}|$ for the four exchange-rate series.

Figure 6 shows clear volatility clustering in all four series. The largest shared spike occurs in March 2020 during COVID, followed by a sustained period of elevated volatility, with smaller clusters appearing again around 2022-23. Large movements are followed by further large movements, while calm periods tend to persist, which is the standard pattern of time-varying conditional variance and strongly supports the use of ARCH/GARCH models.

The plots also hint at fat-tailed behaviour, since large spikes appear relatively often, although this cannot be confirmed from visual evidence alone. Because the series are shown in absolute values, leverage effects cannot be assessed. Overall, the data appear broadly stationary in mean but not in variance.

Question 6

The absolute return plots in Question 5 show clear volatility clustering, suggesting that the exchange-rate returns are not conditionally homoskedastic. To test this formally, I applied the Breusch-Pagan LM test for ARCH effects to the residuals from preliminary ARMA models for each return series. The null hypothesis of no ARCH effects is strongly rejected for all four series, with p-values of 5.7×10^{-60} for CNY, 1.5×10^{-46} for USD, 5.0×10^{-76} for TWI, and 3.9×10^{-36} for SDR. This provides strong evidence that the conditional variance changes over time, so an ARMA-GARCH framework is appropriate.

For each return series, I then estimated $\text{ARMA}(p, q)\text{-GARCH}(p_\sigma, q_\sigma)$ models with $p, q \in \{0, 1, 2\}$ and GARCH orders $(1, 1)$, $(1, 2)$, $(2, 1)$ and $(2, 2)$, using Gaussian innovations. Models were ranked by AIC, with BIC used as a secondary criterion. Adequacy was assessed using the autocorrelation functions of the standardised residuals and squared standardised residuals. For the selected models, residual autocorrelations are small, and the squared standardised residuals show no persistent remaining ARCH pattern. The following models were chosen:

Series	Specification	AIC	BIC	$\max \hat{\rho}_{z^2} $
CNY	ARMA(2,2)-GARCH(1,1)	1.5548	1.5771	0.056
USD	ARMA(2,2)-GARCH(2,2)	1.8352	1.8630	0.078
TWI	ARMA(2,2)-GARCH(2,2)	1.3109	1.3387	0.063
SDR	ARMA(2,2)-GARCH(1,2)	1.8204	1.8455	0.063

The diagnostics support the adequacy of these specifications. The standardised residuals show little evidence of remaining linear dependence, indicating that the ARMA components capture the conditional mean reasonably well. The squared standardised residuals also show no systematic autocorrelation pattern, suggesting that the GARCH components have captured most of the time-varying volatility. The persistence estimates are all below one, so the fitted conditional variance processes are covariance-stationary, although USD and SDR are close to the IGARCH boundary and therefore show highly persistent volatility.

Figure 7 plots the estimated conditional standard deviations from the selected models. All four series show a large volatility spike in March 2020, consistent with the COVID shock, followed by smaller volatility clusters around 2022-23. After 2023, estimated volatility generally moves back toward its long-run level, although CNY remains slightly elevated into early 2026. The estimated volatility patterns are therefore consistent with the evidence from Question 5: exchange-rate returns are approximately stationary in mean, but their conditional variance changes substantially over time.

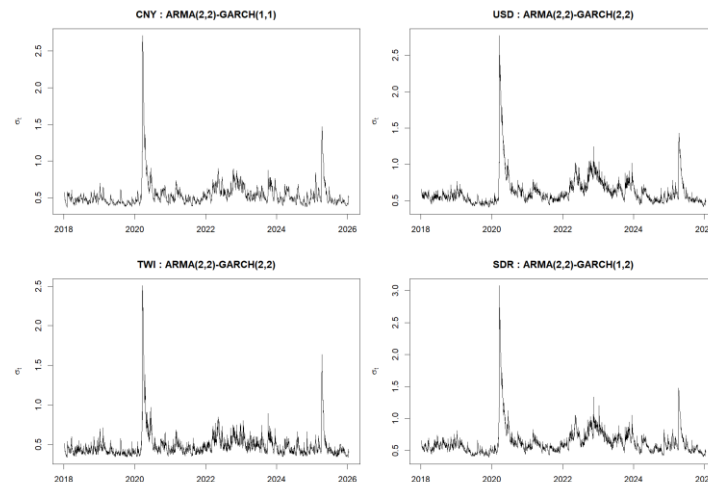


Figure 7: Estimated conditional standard deviation σ_t from each chosen ARMA-GARCH model.

Question 7

For a covariance-stationary GARCH model, the unconditional variance is:

$$\sigma^2 = \frac{\omega}{1 - \sum \alpha_i - \sum \beta_j}$$

where $\sum \alpha_i + \sum \beta_j$ measures volatility persistence. A finite unconditional variance exists only when this persistence term is strictly below one. If persistence is equal to or greater than one, the conditional variance does not converge to a stable long-run level, so the unconditional variance is not defined. In this case, all four selected models have persistence below one, ranging from 0.903 for TWI to 0.962 for SDR. Therefore, a model-implied unconditional variance can be computed for each series.

Series	Persistence	σ_{model}^2	σ_{sample}^2	diff
CNY	0.943	0.305	0.337	0.032
USD	0.961	0.394	0.438	0.044
TWI	0.903	0.239	0.268	0.028
SDR	0.962	0.439	0.449	0.010

The model-implied variances are close to the sample variances from Question 5, suggesting that the selected GARCH models provide reasonable long-run volatility estimates. The differences are small, with SDR showing the closest match and TWI the largest gap. Most model-implied variances are slightly below their sample counterparts, although USD is slightly higher. This is not surprising, since the sample variances are influenced by large volatility episodes such as the March 2020 COVID shock, while the GARCH estimates represent the long-run variance implied by the fitted model.

The persistence estimates are high for all four series, especially USD and SDR, which are closest to the unit-persistence boundary. This means volatility shocks decay slowly, consistent with the volatility clustering observed earlier. However, because persistence remains below one in every case, the fitted variance processes are still stable and have finite unconditional variances.

Question 8

For each selected ARMA-GARCH model, I computed one- and two-step-ahead forecasts of the conditional mean and variance for each return series. Assuming normally distributed innovations, the forecast return is modelled as:

$$r_{t+h} \sim N(\mu_{hat,t+h}, \sigma_{hat,t+h}^2)$$

Since returns are already measured in percentage terms, the 0.01% threshold is entered as 0.01. The probability of a return below this threshold is then calculated by standardising the threshold and applying the standard normal CDF. The resulting probabilities are reported below:

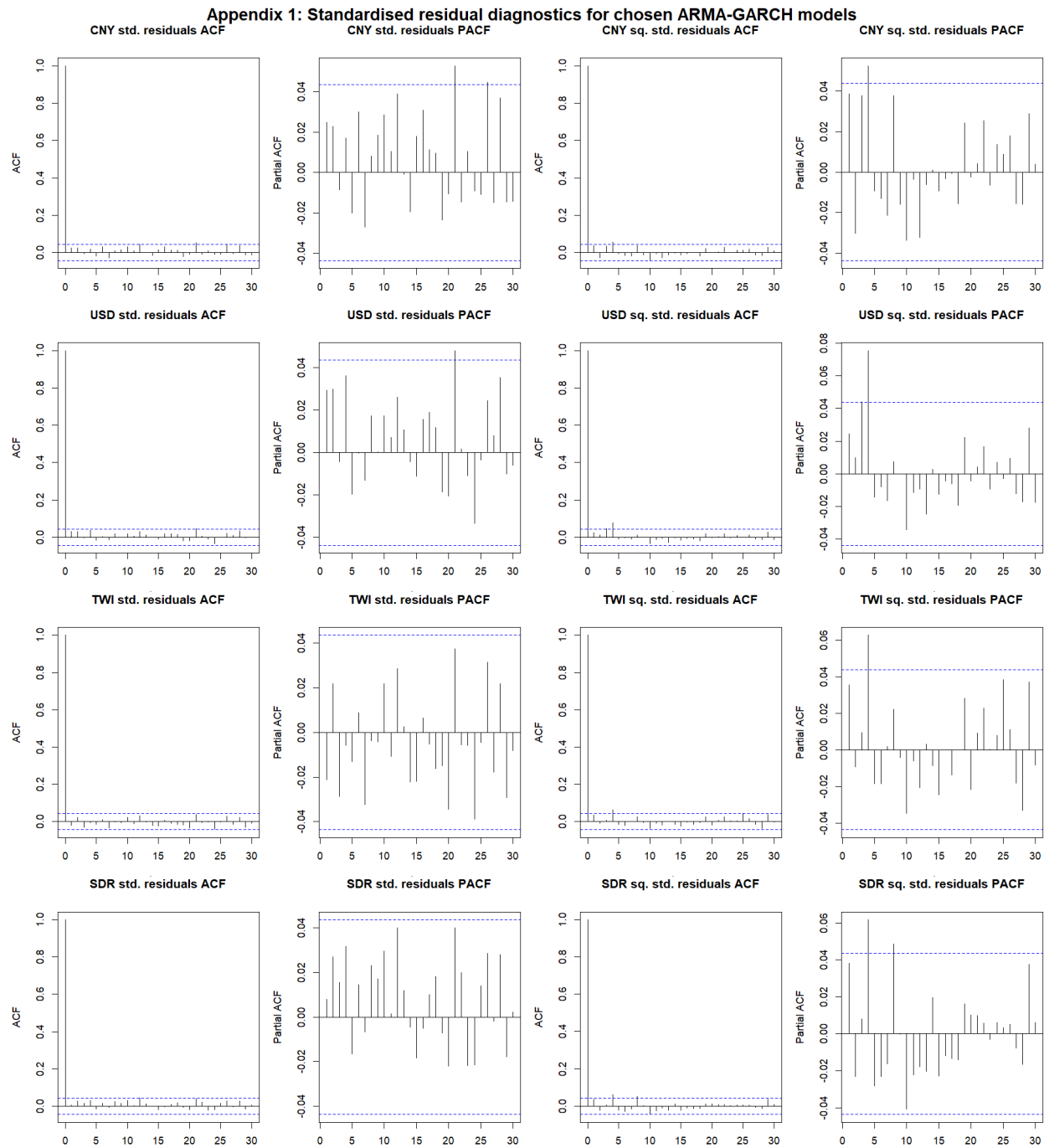
Series	Day	μ^2	σ^2	$P(\text{return} < 0.01)$
CNY	2026-01-13	0.0078	0.4887	0.5018
CNY	2026-01-14	0.0074	0.4926	0.5021
USD	2026-01-13	-0.0675	0.4920	0.5626
USD	2026-01-14	-0.0654	0.4901	0.5611
TWI	2026-01-13	-0.0543	0.4005	0.5638
TWI	2026-01-14	-0.0014	0.4075	0.5111
SDR	2026-01-13	-0.0480	0.4753	0.5486
SDR	2026-01-14	-0.0349	0.4804	0.5372

The probabilities are all close to 0.5 because the 0.01% threshold is very close to the forecast mean for each series. This means the exercise is essentially measuring the short-run probability of a very small or negative return. From a risk-management perspective, the preferred currency is the one with the lowest probability of falling below the threshold, since it has the lowest forecast downside risk.

On 13 January 2026, CNY has the lowest probability of a return below 0.01% at 0.5018, followed by SDR, USD and TWI. On 14 January 2026, CNY again has the lowest probability at 0.5021, followed by TWI, SDR and USD. Therefore, based on this downside-risk criterion, CNY would be the preferred investment on both days. It also has the highest forecast mean return in both periods, which strengthens this conclusion.

However, the differences between the probabilities are small, so the result should not be interpreted too strongly. The threshold is also very close to zero, so it captures only mild downside risk rather than a severe loss event. In practice, I would use these probabilities as one short-horizon risk indicator, alongside other measures such as forecast volatility, VaR or expected shortfall. Still, using the specific criterion in the question, CNY is the most attractive currency because it has the lowest probability of underperforming the 0.01% return threshold on both forecast dates.

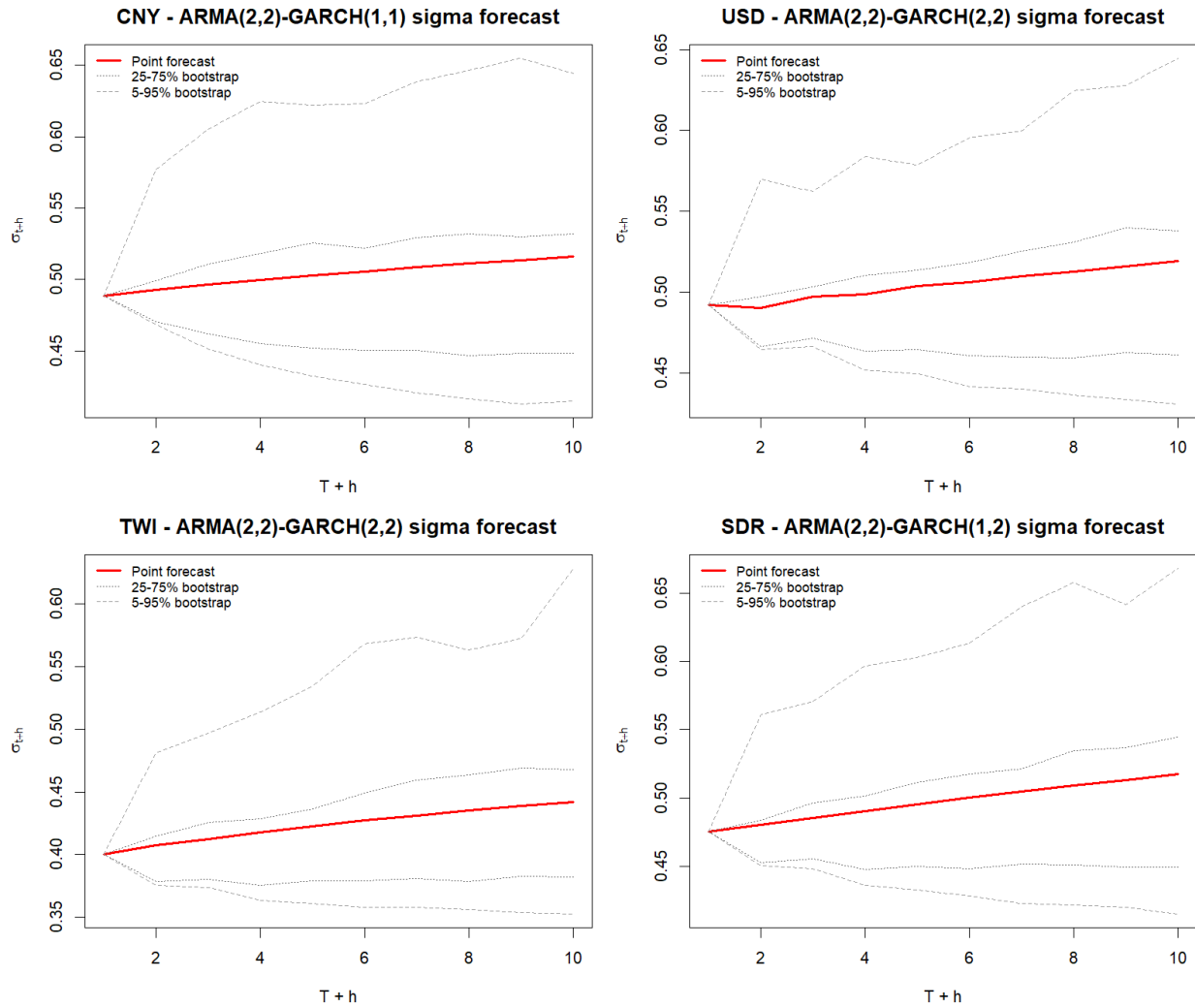
Appendix 1: ARMA-GARCH residual diagnostics



Appendix 1: Standardised residual and squared standardised residual ACF/PACF for each series.

Appendix 2: 10-day sigma forecasts with bootstrap bands

Appendix 2: 10-day sigma forecasts with bootstrap error bands



Appendix 2: 10-day sigma forecasts with 5-95% and 25-75% bootstrap bands.