

Task 1

The Phillips curve implies a stable negative relationship between inflation and unemployment, where higher unemployment should be associated with lower inflation. In this task, I test whether this relationship is supported in the Australian data using stationarity tests and an ARDL model.

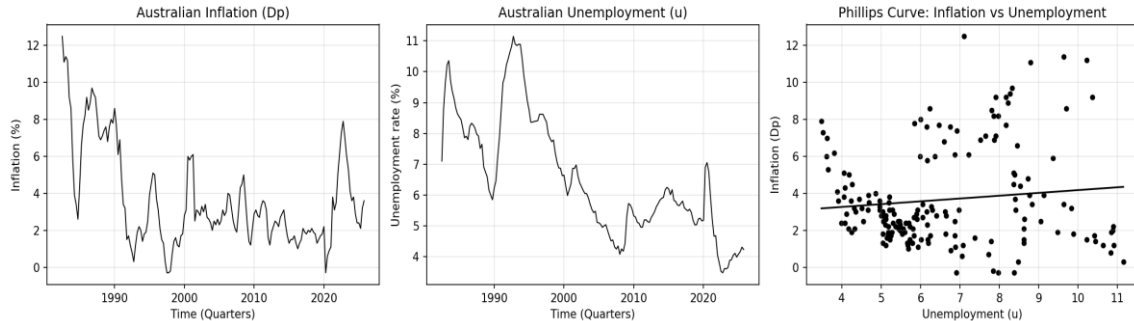


Figure 1. Inflation, unemployment and the static Phillips curve scatterplot

Visually, figure 1 suggests that a simple static Phillips curve is unlikely to describe the data well, with inflation appearing persistent and shifting across regimes, particularly between the high-inflation period of the 1980s and the lower-inflation period that followed. Unemployment, on the other hand, is smoother and moves in longer cycles. The scatterplot also does not show a clear downward-sloping relationship and instead, the fitted line is slightly upward sloping, which conflicts with the original Phillips curve prediction.

Augmented Dickey-Fuller (ADF) tests were then estimated for both inflation and unemployment in levels and first differences, using no-constant, drift-only and drift-plus-trend specifications, with lag lengths from 0 to 9. The ADF auxiliary regressions were compared using AIC/BIC and residual diagnostics. For inflation, the preferred level specification was an ARIMA(8,0,0)-type ADF regression, and the broader set of adequate specifications generally rejected the unit-root null. This supports treating inflation as stationary in levels, which is a reasonable assumption because Dp is already measured as a transformed inflation rate.

The results for unemployment were weaker where across the main level specifications, the ADF tests generally failed to reject the unit-root null, while first differences were clearly stationary. This suggests that unemployment is closer to an $I(1)$ process in this sample. This does not necessarily mean unemployment is a exact random walk with a more plausible interpretation being that the sample contains slow structural shifts in the labour market, including changes in the natural rate of unemployment. However, the result still matters because the mismatch between $I(0)$ inflation and near- $I(1)$ unemployment weakens the case for a stable long-run Phillips curve.

As a benchmark, a static Phillips curve was estimated:

$$Dp_t = \alpha + \beta u_t + \varepsilon_t$$

The results provide little support for the simple static relationship. The estimated coefficient on unemployment was positive rather than negative ($\hat{\beta} = 0.152$), statistically insignificant ($p = 0.147$), and the model explained very little variation in inflation ($R^2 = 0.012$). The residuals also showed strong autocorrelation, suggesting that the static regression omits important dynamic adjustment which motivates moving beyond a simple OLS regression to a dynamic specification.

Given the weak static relationship and the persistence in both series, I estimated ARDL(p,q) models with inflation as the dependent variable and unemployment as the explanatory variable. Models were compared using AIC and BIC, with stability checked through the roots of the lag polynomial. ARDL(8,1) was selected

ECON3350: Applied Econometrics for Macroeconomics and Finance
Research Report 2

because it was stable, minimised BIC, and avoided the overfitting risk of the higher-lag AIC-preferred specifications.

Model	AIC	BIC	F-statistic	F-test p-value
ARDL(8,1)	348.26	385.47	184.4	<0.001

The key results from the selected ARDL(8,1) model are summarised below:

Result	Estimate	p-value	Interpretation
u_t	-0.979	<0.001	Short-run negative Phillips effect
u_{t-1}	0.967	<0.001	Initial effect mostly reverses
Long-run multiplier	-0.253	0.70	Correct sign, but insignificant
Error-correction term	-0.0478	0.065	Slow adjustment toward equilibrium

The ARDL results suggest that the Phillips curve relationship is mainly short-run. The contemporaneous unemployment coefficient was negative and highly significant ($-0.979, p < 0.001$), implying that higher unemployment is associated with lower inflation in the same quarter. However, the one-quarter lag of unemployment was positive and almost the same size ($0.967, p < 0.001$), meaning that much of the initial effect is reversed in the following quarter which explains why the long-run relationship is much weaker than the short-run response.

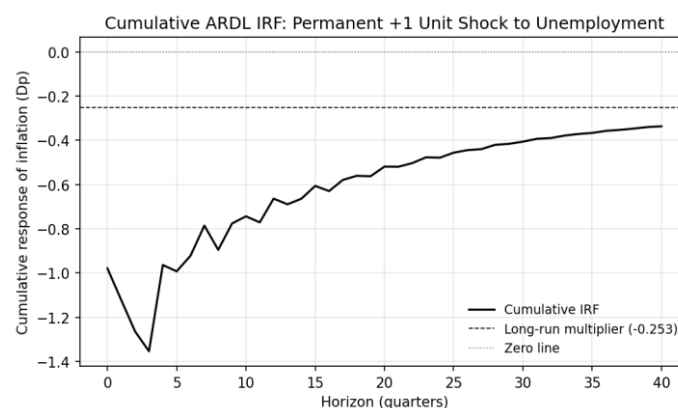
Re-parameterising the ARDL model into its long-run form gives:

$$Dp_t = \mu + \beta u_t + z_t$$

The estimated long-run intercept is $\hat{\mu} = 4.37$, while the long-run unemployment coefficient is $\hat{\beta} = -0.253$. The sign is consistent with Phillips curve theory: a one percentage point increase in unemployment is associated with about a 0.25 percentage point fall in long-run inflation. However, the coefficient is statistically insignificant ($p = 0.70$), so the evidence for a stable long-run Phillips curve is weak.

The error-correction form gives a similar conclusion. The adjustment coefficient is negative (-0.0478), meaning that inflation tends to move back toward its implied Phillips curve equilibrium after a deviation. However, the speed of adjustment is very slow, with less than 5% of disequilibrium corrected each quarter, and the term is only borderline significant ($p = 0.065$). This suggests weak evidence of adjustment rather than strong evidence of a stable long-run equilibrium.

A cumulative ARDL IRF was estimated using the ARDL(8,1) model to trace how inflation responds to a permanent one-point increase in unemployment.



ECON3350: Applied Econometrics for Macroeconomics and Finance
Research Report 2

Figure 2. Cumulative ARDL impulse response of inflation to a permanent one-percentage-point increase in unemployment

Figure 2 shows a sharp initial drop of 0.98 points on impact, consistent with the contemporaneous coefficient. The response then unwinds as the autoregressive structure asserts itself, approaching the long-run multiplier of -0.253 only at very long horizons, at a 40-quarter horizon the IRF still sits around -0.34. The IRF visually reinforces the picture of a dynamically important short-run relationship with weak mean reversion in the long run.

Overall, the evidence does not support a simple stable Phillips curve for Australia. The static model gives the wrong sign, explains little variation in inflation, and leaves autocorrelated residuals. The ARDL model provides stronger evidence of a short-run negative relationship between unemployment and inflation, but this effect is mostly reversed after one quarter. The long-run multiplier is small and statistically insignificant, while the error-correction mechanism is slow and only weakly significant. Therefore, the Phillips curve is more useful as a short-run dynamic relationship than as a stable long-run inflation anchor.

Task 2

The Australian economy was modelled using a six-variable vector containing year-on-year inflation, the unemployment rate, the three-month bank-bill rate, the change in the trade-weighted index, the change in the terms of trade and the change in real per capita US GDP:

$$y_t = (\pi_t, u_t, r_t, \tau_t, \delta t, g_t)'$$

The reduced-form VAR was specified with an intercept but no deterministic trend, since most variables in the system are rates or first differences that do not exhibit any persistent drift, while the non-zero means of inflation and unemployment justify keeping the constant. Writing this out, the model is:

$$y_t = \alpha_0 + A_1 y_{t-1} + \dots + A_p y_{t-p} + e_t$$

Before proceeding, it is assumed that ADF testing identified inflation, unemployment and the interest rate as $I(1)$, and a Johansen trace test on this subset returned a cointegration rank of two at the 5% level. A VECM would therefore be more efficient in principle, however the mixed system of stationary and non-stationary series makes such a specification awkward, so this is treated as background context rather than grounds for re-specification.

VAR(p) models were estimated for every lag order from one to twenty so that the candidate set was wide enough to cover both the parsimonious and the highly parameterised end of the spectrum. Final selection then turned on four diagnostic questions: Are the residuals serially uncorrelated? Are the dynamics stable? Is the lag order parsimonious enough to be supported by the information criteria? Do the other variables carry Granger-causal content for the system?

Stability was clean-cut, with maximum companion roots staying strictly below one for every model up to VAR(17) but rising above one for VAR(18) through VAR(20), where VAR(20) reached 1.016. These three explosive models were dropped immediately, while the remaining models were stable but highly persistent with roots between 0.968 and 0.989, reflecting slow shock decay that is a feature of the data rather than something the lag order can fix.

Result	Estimate	p-value
VAR(2)	0.986	Yes
VAR(6)	0.987	Yes
VAR(11)	0.974	Yes
VAR(17)	0.988	Yes
VAR(18)	1.003	No
VAR(19)	1.007	No
VAR(20)	1.016	No

Table 1. Maximum absolute companion roots for selected VAR specifications.

The BG tests then identified VAR(11) as uniquely well-behaved at short lag horizons, where across all 20 specifications and four BG test horizons, VAR(11) was the only model to fail to reject the null of no residual autocorrelation at any horizon, with p-values of 0.384 at lag 1 and 0.188 at lag 2. Every other lag order rejected the null at the 5% level even at the shortest test horizon, while every model including VAR(11) rejected the null at horizons of 4 and 8. This indicates that some predictable structure remains in the residuals at longer lags for every candidate, however VAR(11) is clearly the best of the available models at capturing short-run dynamics.

ECON3350: Applied Econometrics for Macroeconomics and Finance
Research Report 2

VAR model	BG lag 1	BG lag 2	BG lag 4	BG lag 8
VAR(2)	0.003	< 0.001	< 0.001	< 0.001
VAR(6)	0.013	0.007	<0.001	< 0.001
VAR(11)	0.384	0.188	< 0.001	< 0.001
VAR(13)	< 0.001	< 0.001	< 0.001	< 0.001
VAR(17)	< 0.001	< 0.001	< 0.001	< 0.001

Table 2. Multivariate Breusch-Godfrey LM test p-values across selected VAR specifications

Information criteria gave conflicting guidance, with AIC strongly favouring high-order models (VAR(20) at -2.37) and BIC strongly favouring parsimony (VAR(2) at 4.52). This split is expected in a system with six variables where BIC penalises the n^2p coefficient explosion much more heavily than AIC. VAR(11) already contains 67 regressors per equation and 402 coefficients in the full system, which is a large estimation burden in 174 observations. The IC values were used only as a sanity check on the BG and stability filters, where VAR(11)'s AIC of 2.20 is the third-lowest among stable models and confirms it is not absurdly wasteful relative to nearby alternatives.

Granger non-causality was assessed by testing whether each candidate variable Granger-caused the remaining variables in the system. For VAR(2) and VAR(6), every variable except the terms of trade had a statistically significant causal effect on the system. For VAR(11), unemployment, the exchange rate and US GDP remained significant, the interest rate became marginal ($p = 0.129$), and the terms of trade was again insignificant. The interest rate result for VAR(11) reflects the higher-lag model dispersing its predictive content across many lag coefficients which reduces the joint F statistic, however the underlying conclusion is robust where the terms of trade consistently fails to Granger-cause the system.

VAR model	u	r_t	τ_t	δ_t	g_t
VAR(2)	0.001	< 0.001	< 0.001	0.093	0.006
VAR(6)	0.006	0.001	< 0.001	0.967	0.041
VAR(11)	0.002	0.129	0.022	0.968	0.012

Because the companion roots sit close to one across the models, shocks decay slowly, and a VAR with few lags cannot account for this, leaving serial correlation in the residuals. This is what the BG tests show at the lower lag orders. Adding more lags reduces the problem in the short run, however with six variables the number of parameters grows quickly as lags increase, and models above VAR(17) become unstable. This shows that residual adequacy, stability and parsimony cannot all be satisfied at once in this dataset, which is the main difficulty in choosing a lag order and it also explains why there is no point testing lag orders beyond 20, since every model in that range is already explosive.

Combining all four criteria, VAR(11) was selected as the final model. The reasoning is that VAR(11) is the only specification that passes the BG test at any horizon, is stable with a maximum root of 0.974, has a competitive AIC of 2.20, and retains the same broad Granger-causal structure as the lower-order candidates with unemployment, the exchange rate and US GDP all entering significantly. VAR(2) remains a useful parsimonious benchmark because of its minimum BIC and sharper Granger evidence on the interest rate, however its lag 4 and lag 8 BG rejections mean that the residuals retain predictable structure and any inference drawn from its standard errors is likely to be misleading, so its significant causal links may just reflect noise or missing dynamics rather than genuine predictive content.

ECON3350: Applied Econometrics for Macroeconomics and Finance
Research Report 2

The selected VAR(11) is best treated as a useful approximation of the Australian economy rather than a complete structural representation. The persistent companion root close to unity means that shocks decay slowly, while the BG rejections at lag 4 and 8 indicate that some residual structure is not being captured even by this 402-coefficient specification. The high parameter count relative to the sample size of 174 also makes individual coefficients imprecisely estimated, which is part of why the interest rate loses its Granger significance once 11 lags are estimated. These limitations matter most for impulse responses at longer horizons and for forecasting, where parameter uncertainty translates directly into wide forecast intervals.

Several variables that would have been valuable to include are not present in the dataset. The most important omission is a direct measure of Australian aggregate demand such as real GDP per capita or an output gap, because unemployment only measures slack in the labour market and reacts slowly, so it misses much of the movement in output and demand pressure that drives inflation. Wage growth is a second important omission since inflation in Australia has a strong cost-push component through unit labour costs, and without a wage variable the VAR is forced to attribute persistence in inflation to its own lags alone. Inflation expectations would similarly improve the model since actual inflation is partly determined by what firms and households expect, and the absence of an expectations channel forces the VAR to be purely backward-looking.

On the external side, US per capita GDP is included as a global activity proxy, however it is not the most relevant indicator for Australia where China is the dominant trading partner, so a Chinese activity variable or global commodity prices would be more informative. Fiscal policy is also entirely absent from the model where the bank-bill rate captures the monetary stance but there is no fiscal variable, despite government spending and budget balances being substantial drivers of demand in the recent sample. Including all of these variables would help mitigate the residual autocorrelation problems found above, however each additional variable adds 6p parameters per equation, so there is a real trade-off between fidelity and feasibility in a system of this size.

Task 3

The VAR(11) selected in Task 2 was re-estimated on the sample running from 1982Q3 to 2024Q4, which holds the 2025 quarters out for Task 4. Structural impulse responses were then identified by a standard Cholesky decomposition of the residual covariance matrix, which imposes a recursive ordering of contemporaneous shock transmission through a lower-triangular impact matrix.

The ordering used was:

$$(g_t, \delta_t, \pi_t, u_t, r_t, \tau_t)'$$

The two external variables sit at the top of the ordering, since shocks to US real per capita GDP and Australian terms of trade can plausibly affect domestic variables within the same quarter, while domestic shocks should not affect global activity or world commodity prices contemporaneously. The two slow-moving domestic variables, inflation and unemployment, sit next on the standard argument that prices and labour-market quantities adjust gradually. The interest rate is placed after these to allow the central bank to respond to within-quarter inflation and unemployment outcomes when setting the cash rate. The exchange rate is placed last, since it is a forward-looking financial price that can respond contemporaneously to every other shock in the system. This is the standard small-open-economy ordering and matches the structural priors discussed in Dungey and Pagan (2000).

$$\begin{pmatrix} g_t \\ \delta_t \\ \pi_t \\ u_t \\ r_t \\ \tau_t \end{pmatrix} = \begin{pmatrix} * & 0 & 0 & 0 & 0 & 0 \\ ** & 0 & 0 & 0 & 0 & 0 \\ *** & 0 & 0 & 0 & 0 & 0 \\ **** & 0 & 0 & 0 & 0 & 0 \\ ***** & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{g_t} \\ \varepsilon_{\delta_t} \\ \varepsilon_{\pi_t} \\ \varepsilon_{u_t} \\ \varepsilon_{r_t} \\ \varepsilon_{\tau_t} \end{pmatrix}$$

Two shocks were considered. The first is a one-standard-deviation shock to the terms of trade, which produces an immediate increase of 3.33 in δ_t on impact. The second is a contractionary monetary policy shock, where the task specifies a +0.5% movement in the interest rate. The default Cholesky shock to r produced a 0.575 percentage point on-impact response, so the monetary policy IRFs were rescaled by a factor of 0.870 to deliver the requested +0.5% shock cleanly. Bootstrap 95% confidence intervals were constructed from 500 residual bootstrap replications.

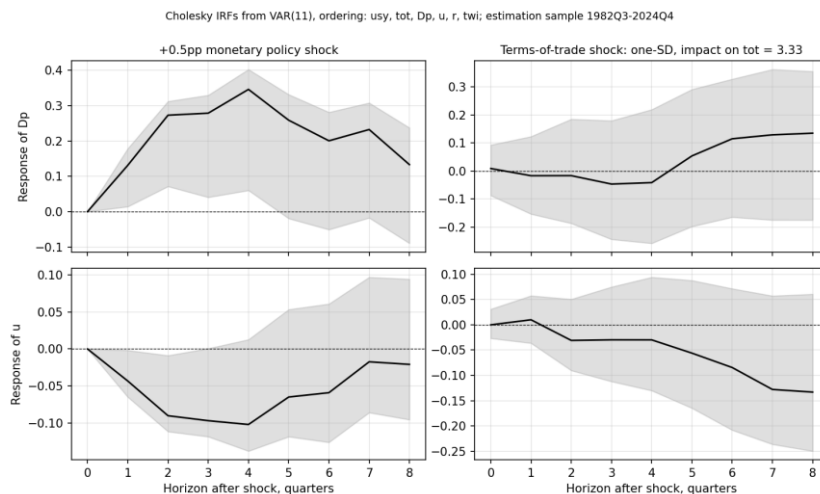


Figure 3. Cholesky IRFs from VAR(11)

For the terms-of-trade shock (right column of Figure 3), inflation barely moves at first and is marginally below zero for the first four quarters, before turning positive from the fifth quarter and reaching about 0.135 two

years out. A lag like this fits the data, since a favourable terms-of-trade movement lifts export income over time and only reaches prices once demand has picked up, rather than passing straight into the price level. Unemployment follows a similar path, sitting near zero early on and then falling further out to about -0.133 by quarter eight, in line with the idea that a better trading position supports activity and tightens the labour market. In both cases the 95% bands include zero at every horizon, so while the point estimates point in a sensible direction, there is little statistical support behind them.

The monetary policy shock (left column of Figure 3) produces a more pronounced and clearly significant short-run response, but with signs that run counter to standard textbook intuition. Inflation and unemployment are both held at zero in the impact period, since they are ordered ahead of r . After that, inflation climbs quickly and reaches a high of 0.346 percentage points above baseline in the fourth quarter, then drifts back down to 0.133 by the end of the second year. Unemployment goes the other way, reaching a low of -0.102 percentage points in the fourth quarter before recovering part of the way. The 95% confidence interval on inflation is entirely positive for horizons one through four, and the confidence interval on unemployment is entirely negative for horizons one through three, so the unexpected signs are not artefacts of imprecise estimation.

Standard contractionary monetary policy should reduce inflation and raise unemployment in the short run, so the observation of the opposite pattern is a textbook example of the price puzzle. The most likely explanation is that the recursive VAR fails to fully separate exogenous policy shocks from the RBA's endogenous reaction to information about future inflation. If the RBA raises rates because it expects higher inflation ahead, the model reads the later rise in inflation as a result of the rate increase, rather than as a result of the expectations that prompted the increase in the first place. The fact that the VAR contains no measure of inflation expectations or forward-looking activity intensifies this issue, since the policy rate then carries the predictive content for inflation that those omitted variables would otherwise have carried.

Several caveats reinforce that this monetary policy IRF should not be treated as evidence on the true causal effect of policy. The VAR(11) selected in Task 2 returned an insignificant Granger-causal F-test for the interest rate on the rest of the system ($p = 0.129$), so the lagged interest rate carries weak predictive content even before structural identification is imposed. Residual autocorrelation at longer test lags also indicates that the system is not fully capturing the dynamics, and the high persistence in the companion roots means that shocks decay slowly and IRFs can drift well past their long-run values. The terms-of-trade shock IRFs are less affected by these problems because the ordering treats δ_t as predetermined relative to domestic variables, but their wide confidence bands still warrant caution.

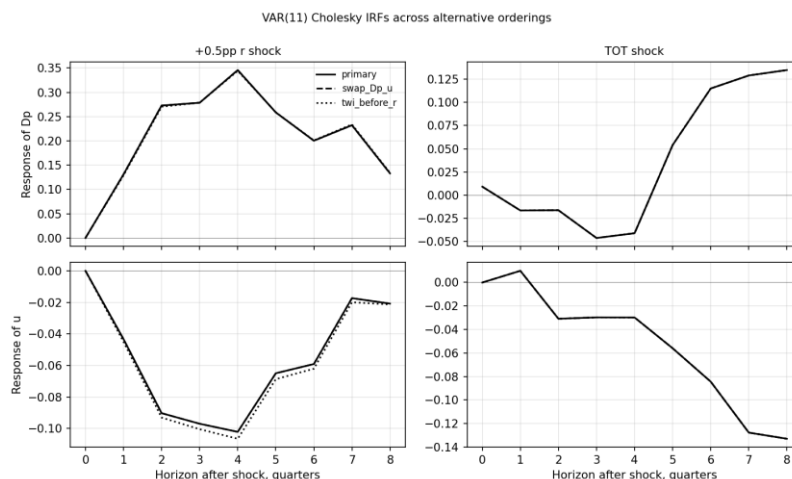


Figure 4. Ordering sensitivity check

ECON3350: Applied Econometrics for Macroeconomics and Finance
Research Report 2

As a robustness check, the IRFs were recomputed under two alternative orderings, namely swapping the positions of inflation and unemployment, and placing the exchange rate ahead of the interest rate. Figure 4 shows that the point responses track each other almost exactly across all four shock-response pairs, which confirms that the main patterns in the responses are not caused by how the domestic variables are ordered. This rules out ordering as the source of the problem, but it does nothing about the price puzzle itself, which comes from leaving out forward-looking and structural variables and not from the ordering.

Taken together, the terms-of-trade shock produces a weak but economically sensible delayed expansion in inflation alongside a gradual fall in unemployment, while the +0.5 percentage point monetary policy shock produces statistically clear but theoretically anomalous responses that almost certainly reflect identification problems rather than the true sign of the policy effect. For the RBA and the Government, the most defensible reading of these results is that the model captures the broad direction of how Australia's external sector transmits to domestic activity, but it cannot be used to draw conclusions about the short-run effect of a deliberate rate change. Any serious policy-relevant identification of monetary policy shocks would require either an expanded VAR that includes forward-looking variables such as market-based inflation expectations, or an alternative identification scheme such as sign restrictions or external instruments that does not rely on a recursive ordering.

Task 4

The VAR(11) selected in Task 2 was re-estimated on a training sample running from 1982Q3 to 2024Q4, with the four quarters of 2025 held out as the evaluation period. Re-estimation is necessary because the original model from Task 2 was fitted using the full sample including 2025, which would mean the inflation forecasts had effectively already 'seen' the realised outcomes. The final training observation is 2024Q4, where headline year-on-year inflation sat at 2.4%, and the holdout values are 2.4%, 2.1%, 3.2% and 3.6% across the four quarters of 2025.

Two diagnostic checks were performed on the re-estimated VAR(11) before generating forecasts. The maximum absolute companion root is 0.9824, which is comfortably below unity and so the model is technically stable, but the proximity to one means the system is highly persistent and shocks decay slowly. Breusch-Godfrey LM tests on the residuals returned a p -value of 0.222 at lag 1 and 0.037 at lag 2, with rejections at lags 4 and 8. The short-run residual behaviour is therefore broadly acceptable on the most powerful (lowest-lag) BG test, although the failures at longer lags indicate the same residual structure problem flagged in Task 2 and motivate cautious interpretation of the forecast intervals.

Four-quarter point forecasts and 95% predictive intervals were then generated from the re-estimated VAR(11). The forecasts, the actual realised values and the two survey-based alternative forecasts are summarised in Table 1.

Quarter	Actual	VAR forecast	VAR 95% PI	Consumer	Market econ.
2025Q1	2.4%	2.39%	1.14% - 3.64%	4.35%	3.20%
2025Q2	2.1%	1.70%	-0.21% - 3.60%	4.43%	3.10%
2025Q3	3.2%	1.68%	-0.80% - 4.17%	4.43%	2.90%
2025Q4	3.6%	2.02%	-1.00% - 5.04%	4.22%	2.70%

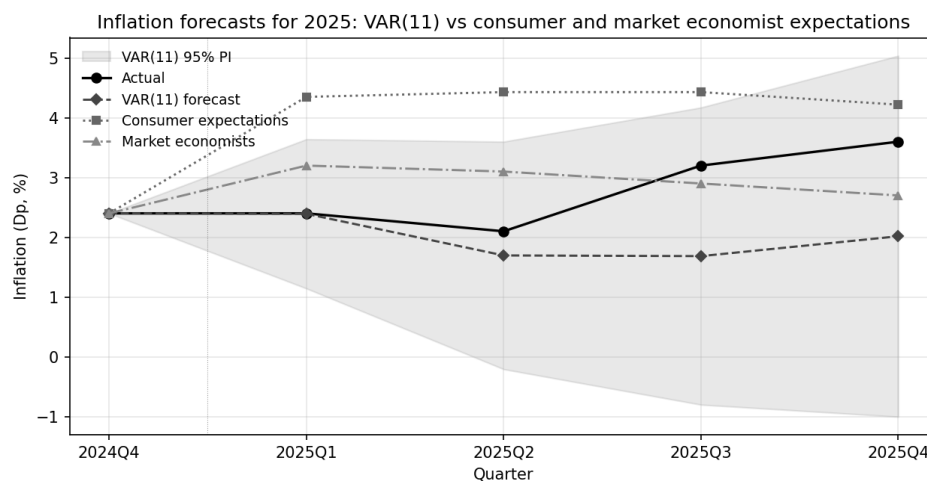


Figure 5. Inflation outcomes and forecasts for 2025

The VAR did best at the start of 2025, where its point forecast for 2025Q1 came in right on the realised 2.4% and its 2025Q2 forecast sat only about 0.4 percentage points under the actual 2.1%. The model performed substantially worse in the second half of the year, missing the rebound to 3.2% in 2025Q3 and to 3.6% in 2025Q4 by roughly 1.5 percentage points each. This pattern is consistent with the model's high persistence, since with a dominant companion root close to one and no mechanism for incorporating new external information such as the 2025 tariff announcements or energy price shocks, the forecast path drifts smoothly

from the last observed inflation rate without anticipating regime changes. The realised figures did stay within the VAR's 95% predictive interval in all four quarters, so on that basis the holdout data does not reject the model, but the intervals get very wide further out, so passing this check does not say much.

Consumer expectations performed poorly throughout the period. The survey responses sat near 4.4% across all four quarters and produced systematic overestimation by roughly 2 percentage points in the first half of 2025, narrowing only modestly as actual inflation rose later in the year. This is the standard finding from the consumer survey literature, namely that household inflation expectations are sticky and tend to anchor to recent salient experiences such as the post-2022 inflation surge, with limited responsiveness to subsequent disinflation. Market economists' forecasts moved in the opposite direction, declining from 3.20% at 2025Q1 to 2.70% by 2025Q4, and were generally close to the realised values across the full period.

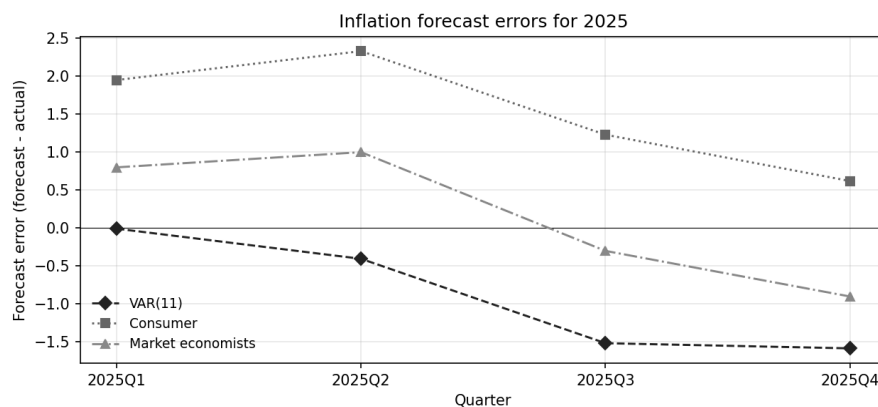


Figure 6. Forecast errors for 2025 inflation, defined as forecast minus actual

Aggregate performance metrics are reported in Table 2. The market economists' forecasts have the smallest mean absolute error (0.75) and RMSE (0.80), the VAR(11) is second on both measures, and the consumer survey is decisively last. The ranking by RMSE is the most useful single summary because it penalises larger errors more heavily, and on that criterion the market economists outperform the VAR by roughly 30%, while the VAR in turn outperforms consumers by a similar margin. The mean error column reveals the direction of the misses, where market economists were almost unbiased on average (mean error 0.15), the VAR underestimated by an average of 0.88 percentage points, and consumers overestimated by 1.53 percentage points.

Approach	Mean error	MAE	RMSE	Max abs. error
Market economists	0.15	0.75	0.80	1.00
VAR(11)	-0.88	0.88	1.11	1.58
Consumers	1.53	1.53	1.67	2.33

Looking quarter by quarter changes the picture slightly. The VAR was closest in 2025Q1 and 2025Q2, where its slow-moving structure happened to match the continued disinflation; the economists were closest in 2025Q3; and the consumers were closest in 2025Q4, mainly because they had held high expectations the whole time and inflation finally rose to meet them. This does not change the overall ranking though, since the VAR gave back its early lead with bigger misses in the second half, and the one quarter the consumers got right was far outweighed by their three large errors before it.

The broader takeaway is that each forecasting approach has a characteristic failure mode that explains its performance. The VAR(11) is internally consistent and uses only data that was available at the time of forecasting, but its mechanical structure and high persistence mean it cannot accommodate news shocks

ECON3350: Applied Econometrics for Macroeconomics and Finance
Research Report 2

such as the 2025 tariff announcements that drove the second-half rebound in inflation. Consumer expectations are anchored to recent salient experiences and adjust slowly, which produces a long-running bias when the underlying environment changes. Market economists combine quantitative models with their own judgement about new information as it arrives, which gives them both the structure that consumer expectations lack and the forward-looking adjustment that the VAR lacks. No single approach has uniformly small errors, but market economists came closest because their methodology is the most flexible.