Financial Econometrics | Lecture 8

Time Series and Forecasting: What Can and Cannot Be Predicted

In Class Group Activity

· SJTU, Summer School 2026 · Fatih Kansoy

Lesson of the day. Some series carry their past forward and can be forecast (sales growth here, $\phi = 0.60$); others have thrown their past away and cannot (stock returns, ϕ near zero). Telling the two apart is half of this lecture. The other half is honesty about a forecast you have made: report an interval, watch for breaks, and trust only tests run on data the model never saw.

Phase 1 – Concept checks

ANSWERS

ConcepTest 1. **Answer.** Option **(d)**. The finding is "returns are unforecastable", not "the model failed"; efficiency predicts exactly this near-zero slope and $\overline{R}^2$. Option (a) chases fit that should not be there; on these data adding lags buys essentially nothing (the AR(2) version has $\overline{R}^2 = 0.0014$) and by AR(4) the $\overline{R}^2$ has turned negative (-0.0022). Option (b) has efficiency backwards; efficiency implies *un*predictable excess returns, because any predictable part is bought now and vanishes before the month begins. Option (c) treats a statistically zero slope ($t = 0.98 < 1.96$) as a tradable signal.

ConcepTest 2. **Answer.** Option **(a)**. QLR is the max of many Chow F's, and a maximum over a search is larger under the null than any single test, so its critical values must be larger too; $6.39 > 6.02$ rejects stability at 1%, with the break located near 1980:Q4. Option (b) ignores selection; comparing a searched maximum with a single-test cutoff overstates the evidence, which is exactly why the QLR tables exist. Option (c) confuses the two routes to nonstationarity; a break is a changed forecasting relationship, a unit root is accumulated permanent shocks, and a QLR rejection says nothing about differencing. Option (d) confuses parameter instability with an outlier; a pooled equation averages two regimes and describes neither, and deleting one quarter does not repair changed coefficients.

A break is a changed relationship, not a unit root; the QLR hunts the first, the Dickey-Fuller test (Lecture 10) hunts the second.

ConcepTest 3. **Answer.** Option **(c)**. Granger causality means "helps predict" and nothing more (the umbrella Granger-causes the rain); the curve arrives earlier, it does not drive the economy. Option (a) reads surviving the autoregressive control as causal, but a third factor that moves both the spread and future growth survives that control too. Option (b) swings too far the other way: an imperfect predictor can still cut forecast errors, and that is all Granger tests. Option (d) thinks a small p rules out confounding; significance says the association is precisely measured, not that it is causal.

Phase 2 – The desk problem

WORKED SOLUTIONS

(a) $\widehat{g}_{T+1} = 1.20 + 0.60(4.00) = 1.20 + 2.40 = \mathbf{3.60}\%$. It sits *below* the last value 4.00, pulled towards the long-run mean because $\phi = 0.60 < 1$. The regression estimates the conditional mean $E(g_{T+1} \mid g_T)$, and among all forecasts the conditional mean has the smallest mean squared forecast error, so no other function of the same history can beat it on average.

(b) Long-run mean $= \dfrac{c}{1-\phi} = \dfrac{1.20}{1-0.60} = \dfrac{1.20}{0.40} = \mathbf{3.00}\%$. Half-life $= \dfrac{\ln 0.5}{\ln 0.60} = \dfrac{-0.6931}{-0.5108} = \mathbf{1.36}$ months (check: $0.60^{1.36} \approx 0.50$). A shock is about half gone within a month and a half: modest memory. On the deck's persistence scale ($\rho_1 = 0.34$ modest, 0.8 high) this series sits between the two.

(c) Interval $= 3.60 \pm 1.96(2.00) = 3.60 \pm 3.92 = [\mathbf{-0.32}\%, \mathbf{7.52}\%]$. It is wide because the RMSFE has two parts: the unknowable future shock (σ_u, captured by the SER) plus the error from estimating c and ϕ, which shrinks like $1/T$. With $T = 240$ the second part is small, so RMSFE $\approx SER$ and the SER captures almost all of the miss; the interval is honest but slightly too *narrow* because it leaves out that estimation part. Unlike a confidence interval, a forecast interval must carry the future shock, so it also needs that shock to be roughly normal.

(d) BIC is smallest at **AR(3)** (507.0); AIC is smallest at **AR(4)** (491.6). BIC charges $\ln(240) \approx 5.5$ per coefficient where AIC charges 2, about 2.7 times more ($5.48/2 = 2.74$); because BIC's price grows with T, a spurious lag cannot beat it forever and BIC settles on the true lag length, while AIC's flat price of 2 lets a spurious extra lag through about 16% of the time even in very large samples. (The table above is on the $-2\log L$ plus penalty scale; the lecture writes the same two penalties divided by T, and the ratio between them, about 2.7, is identical.) On the real lab inflation series, BIC picks AR(3) and AIC keeps AR(6), the same pattern.

(e) $\widehat{r}_{T+1} = 0.50 + 0.03(6.00) = 0.50 + 0.18 = \mathbf{0.68}\%$, barely different from the unconditional mean of about 0.52% ($0.50/0.97 = 0.5155$, rounds to 0.52). With $\phi \approx 0$ the conditional mean collapses to the unconditional mean, so the best forecast *is* the mean. That is the efficient-markets result, and it is the same verdict the lecture reaches on the CRSP index (slope 0.050, SE 0.051): an unforecastable return is the signature of a competitive market where any tradable pattern would already have been arbitraged away.

Phase 3 – The case

REVEAL AND ANSWERS

---

1. Two reasons the in-sample numbers mislead: (i) the dependent variable is a 12-month *overlapping* return, so the errors are serially correlated and ordinary standard errors are far too small, so they overstate significance; (ii) the predictor is close to a random walk, and the overlapping 12-month return is itself so persistent that it behaves like one in this sample, so regressing one on the other can manufacture a high R^2 and a large t by chance: the spurious-regression trap.

2. The point estimate (2.4) does *not* change; only its honest uncertainty grows. With HAC (Newey-West) standard errors at 12 lags, the t falls from 4.8 to about 1.5, so the predictor is no longer significant. It matters because the subscriber was sold a "significant" signal that was an artefact of the wrong standard errors.

3. An in-sample record is almost worthless because in-sample fit always improves as you tune a model, and a spurious fit looks perfect in the very sample it was fitted to. The one test that settles it is a *pseudo-out-of-sample* forecast: estimate using data through some date s only; forecast $s+1$; record the error; move s forward and repeat. Then compare the resulting RMSFE against simple benchmarks (the historical mean, always-zero, an AR forecast). A useful model must *beat* the benchmark, not merely achieve small errors.

4. **Instructor reveal.** The junior analyst was right on all three counts. With HAC standard errors the t fell to 1.5; both series failed to reject a unit root (Dickey-Fuller near -1.5, short

of the 5% value -2.86, a Lecture 10 critical value); and the pseudo-out-of-sample R^2 was about -0.05, worse than forecasting the mean. Roche's 41% R^2 was a spurious regression between two highly persistent series, dressed up with ordinary standard errors on overlapping data. Live from 2007, AlphaSignal's calls were no better than chance, and through the 2008 crash they were catastrophically wrong, a textbook forecast breakdown. The consultant declined.

The textbook runs exactly this experiment on the log dividend yield as a predictor of monthly excess returns. In sample, $t = 2.25$, comfortably past 1.96. Estimated on 1960 to 1992 and forecast over 1993 to 2002, the model's RMSFE was 4.08%; simply forecasting zero gave 4.00% and the historical mean gave 3.98%. The significant star lost to doing nothing. Roche's letter is the same story with worse standard errors. The lesson is not that the FTSE is unforecastable in principle; it is that an impressive in-sample fit, built on overlapping data with the wrong standard errors, no unit-root check and no out-of-sample test, is exactly what an unforecastable market looks like when you fool yourself.

Reserve alternate (butter and the S&P). A 1990s analyst found Bangladeshi butter production "predicted" the S&P 500 with $R^2 = 0.75$. Two unrelated trending series produce a high R^2 by chance; a Dickey-Fuller test and differencing before regressing levels is the defence; out of sample the relationship vanishes.

End of worked solutions.