Financial Econometrics | Lecture 8

Time Series and Forecasting: What Can and Cannot Be Predicted

In Class Group Activity

· SJTU, Summer School 2026 · Fatih Kansoy

How this block runs. You work in your home group of three or four, textbook open. Roles rotate: **Scribe**, **Reporter**, **Sceptic**, **Timekeeper**. All money today is in US dollars ($), pounds sterling (£) and euros (€).

- **Phase 1, Concept checks (~10 min).** Vote, talk for 90 seconds, vote again.
- **Phase 2, Desk problem (~20 min).** Forecast with an AR(1), with an interval.
- **Phase 3, Case (~25 min).** One real story, ending hidden.

A short word list. The **autocorrelation function (ACF)** gives the correlation of a series with its own past at each lag. An **AR(1)** is $x_t = c + \phi x_{t-1} + u_t$ (the lecture writes $\beta_0 + \beta_1 Y_{t-1}$; c and ϕ are the same objects); ϕ measures **persistence** and the long-run mean is $c/(1-\phi)$. The **RMSFE** is the typical forecast miss; it carries the future shock plus estimation error. **BIC** and **AIC** choose the lag length, BIC the more parsimonious (it prefers fewer lags). A series is **stationary** when its mean, variance and lag pattern are stable over time, so the past can teach the future. A **unit root** ($\phi = 1$) makes shocks permanent, and regressions of one such series on another can look strong by chance (**spurious regression**). A **structural break** is a changed forecasting relationship; the **QLR** test searches for it. A **pseudo-out-of-sample** test judges a model on data it never used. **Granger** prediction is not causation. The **efficient-markets** result is that excess returns are unforecastable, so a near-zero R^2 there is the *correct* answer. Overlapping or serially correlated data need **HAC** standard errors (Lecture 7).

Phase 1 – Concept checks

~10 MIN

ConcepTest 1. You fit an AR(1) to monthly excess returns on the CRSP value-weighted US stock index, 1960 to 2002 ($T = 516$). The slope is 0.050 with standard error 0.051 (so $t \approx 0.98$) and $\overline{R}^2 = 0.0006$. A classmate says "this model is useless, throw it out." The best reading of the output is:

(a) The model is specified too short: one lag cannot capture momentum that builds over several months, so the fix is to add lags until the fit statistics start to improve.

(b) The efficient-markets theory fails here: an efficient market should reward investors with returns that follow a stable, predictable pattern, and this output shows no such pattern.

(c) The slope is positive, so last month's return does carry information about this month's; a disciplined trader could exploit it once trading costs are brought low enough.

(d) The output is informative as it stands: an insignificant slope and a tiny $\overline{R}^2$ are just what an efficient market produces, because predictable excess returns get traded away.

ConcepTest 2. For an ADL model of quarterly GDP growth on the term spread ($q = 3$ coefficients that may change), you compute the Chow F statistic at every candidate break date in the central part of the sample. The largest value is $F = 6.39$, at 1980:Q4. The QLR critical values are 4.71 (at 5%) and 6.02 (at 1%); the ordinary F cutoff at 5% would be about 2.60. The right

conclusion is:

(a) There is evidence of a break at the 1% level: a date picked by searching over many candidates must be judged against the larger QLR values, and 6.39 clears 6.02.

(b) Each $F(\tau)$ is an ordinary Chow statistic, so the honest comparison is with the usual cutoff of about 2.60; the special QLR values make the same rejection look weaker than it is.

(c) A rejection this strong means the growth series has a unit root: shocks to it are permanent, so it should be differenced once more before any forecasting equation is refitted.

(d) The best response is to treat 1980:Q4 as an outlier: drop that single quarter, re-estimate on the rest, and keep one pooled equation for the whole of the remaining sample.

ConcepTest 3. In an ADL(2,2) model of quarterly GDP growth, the F statistic on both lags of the term spread is 4.43 ($p = 0.01$), and an inverted yield curve has come before almost every US recession since 1960. A newspaper headline says "the bond market causes recessions." The best reading is:

(a) The headline is right: the spread's lags stay significant after controlling for growth's own history, and beating that control is what separates a genuine cause from a spurious correlation.

(b) The headline is wrong for a simpler reason: the curve has also inverted with no recession following, and a predictor that gives false signals is not worth using, whatever the F statistic says.

(c) The headline is wrong, but the number is real: the spread Granger-causes growth, meaning its past improves the forecast beyond growth's own history, not that yields move the economy.

(d) The headline is right in substance: at $p = 0.01$ the link is far too precise to be produced by a third factor, such as expectations of future policy, so a causal reading is warranted.

Phase 2 – The desk problem

~20 MIN

Forecasting a retailer's sales growth

The setting. You have 240 months of a German grocery retailer's annualised like-for-like (same stores) sales growth g_t (%). OLS gives the AR(1)

$$g_t = 1.20 + 0.60\, g_{t-1} + u_t, \qquad t_{\text{slope}} = 11.5,\ R^2 = 0.36,\ SER = 2.00.$$

The latest value is $g_T = 4.00$.

(a) One-step forecast. Compute $\widehat{g}_{T+1}$. Is it above or below the last value? Why? The lecture says a forecast built this way is the best you can do under squared-error loss. Why?

(b) Long-run mean and persistence. Compute the long-run mean $c/(1-\phi)$ and the half-life of a shock, the number of months h at which half the shock survives, $\phi^h = 0.5$, so $h = \ln 0.5 / \ln \phi$. In one sentence, how much memory does this series have?

(c) Forecast interval. Build the 95% one-step interval, $\widehat{g}_{T+1} \pm 1.96\, SER$. Why is it wide? The lecture splits the RMSFE into two parts; which part does the SER capture, and which part does this interval leave out?

(d) Lag length. You refit AR(1) to AR(4):

	AR(1)	AR(2)	AR(3)	AR(4)
BIC	512.0	508.5	507.0	509.0
AIC	505.0	498.0	493.1	491.6

Which lag length does each criterion choose? Why is BIC more parsimonious?

(e) The efficient-markets contrast. You fit the same AR(1) to a stock's monthly returns and get $r_t = 0.50 + 0.03\, r_{t-1} + u_t$, $t = 0.6$, $R^2 = 0.001$, mean return 0.52. If last month was $r_T = 6.00\%$, what is the best forecast of next month's return? Why is that essentially the mean, and why is that *good* news?

Phase 3 – The case

~25 MIN

The AlphaSignal letter: a backtest too good to be true

How to run this. Read the story with your group. The ending is hidden on purpose. Commit a written position before the reveal. Do not search the internet.

The story. It is early 2007. A London quant, Daniel Roche, launches a paid newsletter, **AlphaSignal**. His pitch: he has found a leading indicator that forecasts the FTSE 100. Each month he regresses the FTSE's next-12-month total return on a single predictor he builds from macro data, using monthly data 1985 to 2006 (about 260 months). His headline regression: slope +2.4, ordinary $t = 4.8$, $R^2 = 0.41$ ("explains 41% of the market"), and the fit line tracks the FTSE beautifully. Subscribers pour in at £500 a year. A pension consultant asks his junior analyst to check three things: (1) Roche's predictor is itself a slow-moving, trending macro series that looks like a random walk; (2) the dependent variable is a 12-month-ahead *overlapping* return, and the standard errors are ordinary OLS, not HAC; (3) Roche only ever reports *in-sample* fit, with no out-of-sample test.

The story stops here.

Four discussion questions (answer all four; take a position):

1. Roche advertises $R^2 = 0.41$ and $t = 4.8$. Name **two** reasons these in-sample numbers can be misleading here.
2. The analyst recomputes the t with HAC standard errors and gets 1.5. Does the point estimate (2.4) change? What changes, and why does it matter? (Standard errors under serial correlation are Lecture 7 material; the rest of the case is this week's.)
3. Roche says "twenty years and never wrong." Why is a perfect *in-sample* record almost worthless as evidence? Name the one test that would settle it.
4. **Commit.** Should the consultant recommend AlphaSignal to clients? State your position and your single strongest reason.

End of activity handout. Solutions are distributed after the block.