Financial Econometrics
Shanghai Jiao Tong University
Summer School 2026
Empirical Exercise 14.2

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition, Global Edition)

Data: `Stock_Returns_1931_2002.csv` *with columns* `time` *(year),* `Month`*,* `ExReturn` *(monthly percentage excess return), and* `ln_DivYield` *(100× log dividend yield). Use 1932:M1–2002:M12.*

Empirical Exercise 14.2. Read the boxes "Can You Beat the Market? Part I" and "Can You Beat the Market? Part II" in Chapter 14. These boxes study whether monthly excess returns on a broad stock index can be forecast from past returns and from the dividend yield, and relate the answer to the efficient markets hypothesis. The dataset `Stock_Returns_1931_2002.csv` is an extended version of the data described in those boxes. It contains the columns `time` (year), `Month`, `ExReturn` (the monthly excess return on the CRSP value-weighted index, measured in percentage points per month), and `ln_DivYield` (100 times the logarithm of the dividend yield). In this exercise you will re-estimate the boxes' regressions over the longer sample 1932:M1–2002:M12.

(a) Repeat the calculations reported in Table 14.2, using regressions estimated over the 1932:M1–2002:M12 sample period. That is, estimate AR(1), AR(2) and AR(4) models for the excess return `ExReturn` and report the estimated coefficients, the intercept, the F-statistic (with its p-value) testing that all lag coefficients are zero, and R^2.

(b) Repeat the calculations reported in Table 14.6, using regressions estimated over the 1932:M1–2002:M12 sample period. That is, estimate autoregressive distributed lag models of `ExReturn` that use the dividend yield `ln_DivYield` (in first differences and in levels, following the specifications in the box) as an additional predictor.

(c) Is the variable ln(*dividend yield*), that is `ln_DivYield`, highly persistent? Explain how you reached your conclusion.

(d) Construct pseudo out-of-sample forecasts of excess returns over the 1983:M1–2002:M12 period, using forecasting regressions that begin in 1932:M1. Compare the root mean squared forecast error (RMSFE) of the candidate models with that of a "zero forecast" (always forecasting a zero excess return) and a "constant forecast" (a recursively estimated model that contains only an intercept).

(e) Do the results in (a) through (d) suggest any important changes to the conclusions reached in the boxes? Explain.

Calculations performed using Stata over the 1932:M1–2002:M12 sample.