DATA 201: Time Series Analysis Linear Regression Model Lecture 4: LRM with one independent variable

Lulu Wang

Data Analytics
Dickinson College

1/31/2025

Introduction

The Linear Regression Model

Objectives

- Review the Linear Regression Model (LRM)
- Discuss its estimation in Python
- Explore applications in financial data

I will review the regression model in broad terms and more details can be found in an introductory statistics/econometrics textbook, such as:

- Stock and Watson, *Introduction to Econometrics*, Pearson
- Wooldridge, *Introductory Econometrics*, A Modern Approach, South-Western

Are Financial Returns Predictable?

- This topic has been widely debated with mixed evidence.
- Common predictor variables:
  - Dividend Price (DP) ratio
  - Earnings Price (EP) ratio
  - Volatility
  - Book-to-Market (BM) ratio
  - Interest Rates and Yield Spreads
- Studies often analyze the future equity premium using current predictor values.
- Some research suggests weak predictability, but real-time usefulness remains questionable.

Some Financial Ratios Used for Prediction

1. **Dividend-Price (DP) Ratio:**
   - Measures how much dividend a company pays relative to its stock price.
   - A high DP ratio means the company is giving back more cash to investors, which can be a sign of a stable company.
2. **Earnings-Price (EP) Ratio:**
   - This ratio tells you how much a company earns (profit) for every dollar you invest in its stock.
   - A high EP ratio means the company is making good profits compared to its stock price, which could indicate a good investment.
3. **Book-to-Market (BM) Ratio:**
   - Compares a company's book value to its market value.
   - If BM is high, the stock might be undervalued (a good deal).

4. Volatility: Measures how much a stock's price fluctuates over time.

Why Do Yield Spreads Matter?

- A yield spread is the difference in interest rates between two bonds with different risks or maturities. It is commonly measured as:

  $$\text{Yield Spread of A and B} = \text{Yield of Bond A} - \text{Yield of Bond B}$$

- For example, investors often compare:
  - Corporate Bonds vs. Government Bonds $\rightarrow$ Measures risk in corporate debt.
  - Long-term vs. Short-term Treasury Bonds (yield curve) $\rightarrow$ Indicates future economic expectations.

Are Financial Returns Predictable? (Cont.)

- Goyal and Welch (2008) perform a comprehensive examination of the performance of many predictors and conclude that:
  - there is some evidence of predictability, and not useful (profitable) in real-time.

Are Financial Returns Predictable? (Cont.)

- The dataset from their research, along with an updated version extending to 2015, is publicly available. This dataset includes monthly, quarterly, and annual financial variables. For this analysis, we focus on:
  - **D12:** Total dividends paid
  - **Rfree:** The risk-free rate (approximated by the 3-month Treasury bill rate)
  - **CRSP_SPvw:** Return on the CRSP value-weighted index
  - **Index:** Stock market index level (price level)

Are Financial Returns Predictable? (Cont.)

We compute:

- **Dividend-Price Ratio (DP):** dividend-to-price ratio for the Index in year t:

$$Dividend - PriceRatio(DP) = \frac{Dividends(D12)}{StockIndexPrice(Index)} \times 100$$

- **Equity Premium (ep_crsp):** equity premium (= return minus riskfree rate) on CRSP value weighted Index in year $t+1$:

$$EquityPremium(ep\_crsp) =$$

$$100 \times (IndexReturn(CRSPSPvw) - RiskFreeRate(Rfree))$$

Are Financial Returns Predictable? (Cont.)

- We visualize the relationship between the equity premium and the dividend-price ratio using a scatter plot.

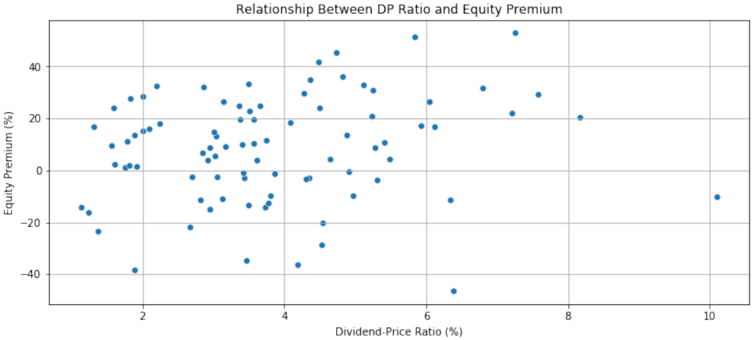

Figure 1: Annual observations of the dividend-price (DP) ratio and the equity premium for the CRSP value weighted Index (calculated as the difference between the next year return on the CRSP Value Weighted Index and the riskfree rate

Are Financial Returns Predictable? (Cont.)

- Scatter plots can sometimes be challenging to interpret when analyzing the relationship between two variables. In such cases, adding a regression line helps visualize the trend more clearly.

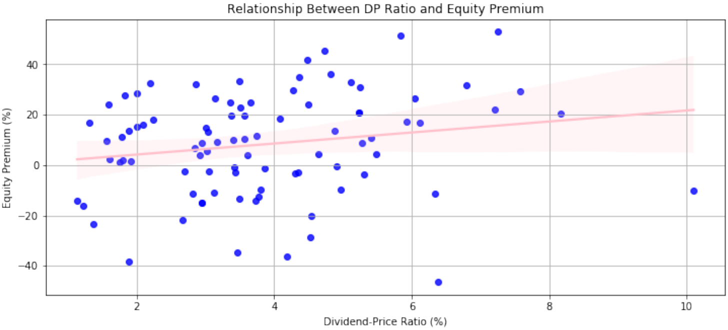

Figure 2: Scatter plot of the Dividend-to-Price ratio and the equity premium with the regression line added to the plot.

Time series of the DP ratio and equity premium

- A scatter plot does not capture time series characteristics that might reveal important patterns. The figures below display the dividend-price (DP) ratio and equity premium over time, beginning in 1926.

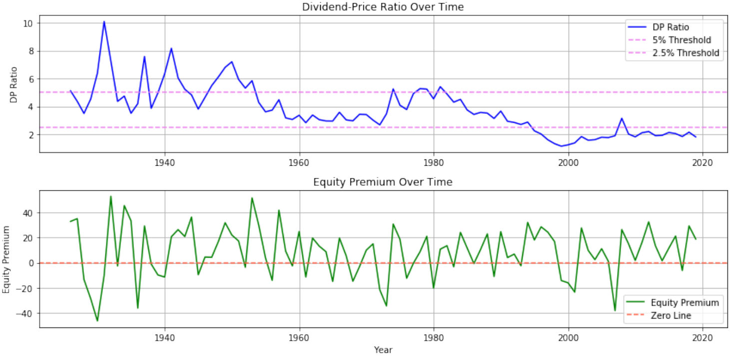

Figure 3: Time series of the DP ratio (top) and the equity premium (bottom) from 1926 to 2015.

The Linear Regression Model (LRM)

- The linear regression model is given by:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t,$$

where:

  - Y_t: dependent variable at time t
  - X_t: independent variable / factor / predictor at time t
  - β_0, β_1: coefficients to be estimated
  - ε_t: error term (mean zero, variance σ^2)

- The **expected** (or average) value of Y_t given X_t is:

$$E(Y_t \mid X_t) = \beta_0 + \beta_1 X_t.$$

Interpretation and CAPM

- **Interpretation**:
  - β_0: the expected value of Y_t when $X_t = 0$
  - β_1: the expected change of Y_t for a unit change of X_t
- The Capital Asset Pricing Model (CAPM) is an example of an LRM:

$$R_t^i \;=\; \beta_0 \;+\; \beta_1 \, R_t^{\mathrm{MKT}} + \varepsilon_t,$$

where R_t^i and R_t^{MKT} represent the excess stock and market returns, respectively.

OLS estimation

- Ordinary Least Squares (OLS) is a method to estimate the coefficients β_0 and β_1 from a sample of observations of X_t and Y_t.
- **OLS recipe**: choose the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the Sum of the Square Residuals (SSR), i.e. $\sum_{t=1}^{T} \hat{\varepsilon}_t^2$.
- Notice that we use ˆ (ˆ) to denote *estimated* quantities of a population parameter (e.g. $\hat{\beta}_1$ vs. β_1).
- In the simple case of the LRM with only one independent variable, we have analytical formulas for the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

OLS formulas for the simple LRM

1. $\hat{\beta}_1 \;=\; \dfrac{\sigma_{X,Y}}{\sigma_X^2} \;=\; \rho_{X,Y}\,\dfrac{\sigma_Y}{\sigma_X}$
   - $\sigma_{X,Y}$: sample covariance of X_t and Y_t
   - σ_X: sample standard deviation of X_t
   - σ_Y: sample standard deviation of Y_t

2. $\hat{\beta}_0 \;=\; \overline{Y} \;-\; \hat{\beta}_1\,\overline{X}$
   - $\overline{Y}$: sample mean of Y_t
   - $\overline{X}$: sample mean of X_t

OLS Example

- Assume that our dependent variable is the return of the equity market in excess of the risk-free rate, and the independent variable is the DP ratio.
- The LRM is then

$$EP_{t+1}^{CRSP} = \beta_0 + \beta_1 DP_t + \varepsilon_t .$$

- The quantity $\beta_0 + \beta_1 DP_t$ represents:
  - the regression line as a function of DP,
  - the expected equity premium in the following time period when the current dividend-price ratio is equal to DP_t, that is,

$$E\left(EP_{t+1}^{CRSP} \mid DP_t\right) = \beta_0 + \beta_1 DP_t .$$

Esimation in Python

- Rather than manually computing these estimates, Python's **statsmodels** automatically calculates regression coefficients, standard errors, and other diagnostic measures:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                ep_crsp   R-squared:                       0.034
Model:                            OLS   Adj. R-squared:                  0.023
Method:                 Least Squares   F-statistic:                     3.061
Date:                Thu, 30 Jan 2025   Prob (F-statistic):             0.0837
Time:                        19:51:46   Log-Likelihood:                -392.57
No. Observations:                  89   AIC:                             789.1
Df Residuals:                      87   BIC:                             794.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1903      5.274     -0.036      0.971     -10.673      10.292
DP             2.1865      1.250      1.749      0.084      -0.298       4.671
==============================================================================
Omnibus:                        3.676   Durbin-Watson:                   1.852
Prob(Omnibus):                  0.159   Jarque-Bera (JB):                3.340
Skew:                          -0.474   Prob(JB):                        0.188
Kurtosis:                       3.021   Cond. No.                         10.9
==============================================================================
```

Fitted Values & Residuals

- Based on the coefficient estimates, we can then calculate the fitted values and the residuals of the regression model.
  - The fitted values:

$$\hat{\mathrm{EP}}_{t+1} = \beta_0 + \beta_1 \times DP_t$$

  - The residuals measure the difference between actual and predicted returns.

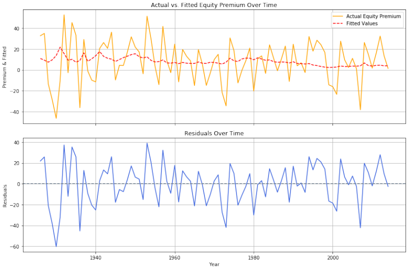

Figure 4: Time series of the realized and predicted equity premium (top) and the residuals obtained as the difference between the realized and predicted equity premium (bottom)