Lesson 2: OLS and GLS for cross-sectional regressions

Some conventions during the slides

- Black font: Basic concepts
- Blue font: More advanced concepts which I encourage you to understand, but not to memorize.
- Red font: Advanced concepts that you can skip if you are not interested in the details.

The cross-section of stock returns

Cross-sectional predictability

- Instead of predicting the return on the aggregate stock market, there is a large literature that studies cross-sectional stock return predictability.
  The typical procedure is:
  - Sort stocks on a characteristic into quintile or decile portfolios and document a pattern in average returns.
  - Construct a long-short strategy that buys the top quintile or decile and shorts the bottom decile.
  - The factor constructed is then used as a factor to explain average returns.

The cross-section of stock returns

- By forming a long-short strategy we "net-out" some of the passive exposures, for instance, to the market that would arise from a long-only portfolio.
- This argument works as long as, for instance, market betas and the characteristics are not highly correlated.
- Equivalently, you can show that a characteristic predicts returns in the cross-section (using the appropriate econometrics.).

How does it (normally) work in practice with U.S. equities

1. Equity prices come from CRSP and accounting data is from Compustat. CRSP-Compustat merged provides a variable PERMNO that you can use to merge the data.
2. Accounting data is released to the public with a lag. It is common practice to lag the accounting data by 6 months (sometimes 3 months) to ensure that the accounting data are indeed available to investors at the time of portfolio formation. Equivalently get the exact date (SEC filing date).

3. Most research uses stocks listed on the AMEX, NASDAQ, and NYSE. Sometimes papers impose a minimum price of $1 or $5 to avoid looking at penny stocks.
4. As most firms have their fiscal year-end in December, it is common practice to sort portfolios in June and then track the performance of the portfolios for the next 12 months.
5. To sort stocks into portfolios, we typically use the characteristics of the NYSE stocks, which tend to be larger firms, to determine which stock goes into which portfolio.
6. Within each of the portfolios, you can either value-weight or equally-weight the stocks. The results are typically stronger for equally-weighted returns as anomalies tend to be more pronounced for smaller stocks. However, value-weighting arguably leads to economically more meaningful results.
7. If a firm defaults, it is important to use the delisting return when available

Basic Equity Return Factors

Main cross-sectional predictors that have been studied in the literature

- Market beta
- Market Capitalization
- Book-to-market ratio
- Lagged price changes
- Investment / asset growth
- Profitability
- Liquidity

Testing for predictive variable in the cross section

Consider the following factor model, assets have common exposure to the factor F and idiosyncratic risk ϵ.

$$R^e_{i,t} = \alpha_i + \sum_{j=1}^{K} \beta_{i,j} F_{j,t} + \epsilon_{i,t+1}$$

What does it mean for factors to be priced?

$$\mathbb{E}[R^e_{i,t}] = \alpha_i + \sum_{j=1}^{K} \beta_{i,j} \lambda_j$$

- What should we expect from an adequate factor model?

$$\mathbb{E}[\alpha_i] = 0 \quad \forall i$$

e.g. no pricing errors, the exposure should be the only source of risk.

Joint Hypothesis Problem

- We need to specify a model for the cross-section of returns and test the model.
- What if the model is misspecified?
- Are assets mispriced because there are arbitrage opportunities or because we are looking at the wrong model?
- **Quick answer**: We can't tell.
- We tend to assume first that a model is correct and then test it.

Quick recap on OLS

Consider the model

$$y = X\beta + \epsilon$$

and

$$\mathbb{E}[\epsilon\epsilon'] = \Omega$$

The OLS estimate is

$$\hat{\beta} = (X'X)^{-1}X'y$$

and the variance is

$$\mathbb{V}[\hat{\beta}] = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Quick recap on OLS

How would you estimate the residuals?

$\hat{\epsilon} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = (I - X(X'X)^{-1}X')y = (I - X(X'X)^{-1}X')(X\beta + \epsilon) = (I - X(X'X)^{-1}X')\epsilon$

so the variance of estimated residuals is

$$\mathbb{V}[\hat{\epsilon}] = (I - X(X'X)^{-1}X')\Omega(I - X(X'X)^{-1}X')'$$

Going back to the factor model

Notation: $\mathbb{E}_T[X] = \frac{1}{T}\sum_{t=1}^{T} X_t$

Consider the model

$$\underbrace{\mathbb{E}_T[R_i^e]}_{N\times 1} = \underbrace{\beta}_{N\times K}\underbrace{\lambda}_{K\times 1} + \underbrace{\alpha}_{N\times 1}$$

where β is the factor exposure matrix, λ is the factor risk premium, and α is the idiosyncratic return. Consider $\mathbb{E}[\alpha\alpha'] = \Omega$

OLS Estimation

$$\hat{\lambda} = (\beta'\beta)^{-1}\beta'\mathbb{E}_T[R_i^e]$$
$$\hat{\alpha} = \mathbb{E}_T[R_i^e] - \beta\hat{\lambda}$$

Variances

$$\mathbb{V}[\hat{\lambda}] = (\beta'\beta)^{-1}\beta'\Omega\beta(\beta'\beta)^{-1}$$
$$\mathbb{V}[\hat{\alpha}] = (I - \beta(\beta'\beta)^{-1}\beta')\Omega(I - \beta(\beta'\beta)^{-1}\beta')'$$

We need to estimate Ω.

Time-series vs. Cross-section

$$R_t^e = a + \beta f_t + \epsilon_t \iff$$
$$\mathbb{E}[R^e] = a + \beta \mathbb{E}[f]$$

Covariance of α_i

$$
\begin{aligned}
\alpha &= \mathbb{E}_T[R^e] - \beta\lambda \\
\mathbb{V}[\alpha] &= \mathbb{V}[\mathbb{E}_T[R^e]] = \mathbb{V}[a + \beta\mathbb{E}_T[f] + \mathbb{E}_T[\epsilon]] \\
&= \mathbb{V}[\mathbb{E}_T[R^e]] = \mathbb{V}[a + \frac{1}{T}\beta\sum f_t + \frac{1}{T}\sum \epsilon_t]
\end{aligned}
$$

Assuming factors and residuals are uncorrelated (the model is well specified).

$$
\begin{aligned}
\mathbb{V}[\alpha] &= \frac{1}{T}\beta\mathbb{V}[f]\beta' + \frac{1}{T}\mathbb{V}[\epsilon] \\
&= \frac{1}{T}\left(\beta\Sigma_f\beta' + \Sigma\right)
\end{aligned}
$$

Altogether

$$
\begin{aligned}
\mathbb{V}[\hat{\lambda}] &= \frac{1}{T}(\beta'\beta)^{-1}\beta'(\beta\Sigma_f\beta' + \Sigma)\beta(\beta'\beta)^{-1} \\
&= \frac{1}{T}\left((\beta'\beta)^{-1}\beta'\beta\Sigma_f\beta' + (\beta'\beta)^{-1}\beta'\Sigma\right)\beta(\beta'\beta)^{-1} \\
&= \frac{1}{T}\left(\Sigma_f + (\beta'\beta)^{-1}\beta'\Sigma\beta(\beta'\beta)^{-1}\right) \text{ eq. 12.12 in Cochrane 2006} \\
\mathbb{V}[\hat{\alpha}] &= \frac{1}{T}\left(I - \beta(\beta'\beta)^{-1}\beta'\right)(\beta\Sigma_f\beta' + \Sigma)\left(I - \beta(\beta'\beta)^{-1}\beta'\right)'
\end{aligned}
$$

Idempotent matrix $I - \beta(\beta'\beta)^{-1}\beta'$

- An idempotent matrix is a matrix that, when multiplied by itself, yields itself. This property is very useful in matrix algebra and econometrics.
- The matrix $I - \beta(\beta'\beta)^{-1}\beta'$ is idempotent because

$$(I - \beta(\beta'\beta)^{-1}\beta')(I - \beta(\beta'\beta)^{-1}\beta') = I - 2\beta(\beta'\beta)^{-1}\beta' + \beta(\beta'\beta)^{-1}\beta'\beta(\beta'\beta)^{-1}\beta' = I - \beta(\beta'\beta)^{-1}\beta'$$

We can call it $M = (I - \beta(\beta'\beta)^{-1}\beta')$. Its transpose is also equal to itself, $M' = M$.

$$M' = (I - \beta(\beta'\beta)^{-1}\beta')' = I' - \beta(\beta'\beta)^{-1'}\beta' = M$$

- $M\beta = \beta - \beta(\beta'\beta)^{-1}\beta'\beta = 0$

Cont.

$$
\begin{aligned}
\mathbb{V}[\hat{\alpha}] &= M(\beta\Sigma_f\beta + \Sigma)M \\
&= M\beta\Sigma_f\beta M + M\Sigma M \\
&= M\Sigma M \\
&= \frac{1}{T}\left(I - \beta(\beta'\beta)^{-1}\beta'\right)\Sigma\left(I - \beta(\beta'\beta)^{-1}\beta'\right)' \text{ eq 12.13 in Cochrane 2006}
\end{aligned}
$$

Testing whether all pricing erros are zero

Use the statistic

$$\hat{\alpha}' cov(\hat{\alpha})^{-1} \hat{\alpha} \sim \chi^2_{N-K}$$

GLS: Generalized Least Squares

- Since the residuals in the cross-sectional regression are correlated with each other, standard textbook advice is to run a GLS cross-sectional regression rather than OLS. (The OLS is not longer the BLUE estimator, e.g. Gauss-Markov theorem does not apply.) More important, the t-statistics are no longer valid.

Quick recap on GLS

Consider the model

$$y = X\beta + \epsilon$$

and

$$\mathbb{E}[\epsilon\epsilon'] = \Omega$$

We are going to estimate a variation of the model such that the residuals are uncorrelated.

Quick recap on GLS

Define

$$\Omega^{-1} = C'C$$

(we can do that with every positive definite matrix Ω and the decomposition is not unique). Let's transform the model

$$Cy = CX\beta + C\epsilon$$

$$\tilde{y} = \tilde{X}\beta + \tilde{\epsilon}$$

$$\mathbb{E}[\tilde{\epsilon}\tilde{\epsilon}'] = \mathbb{E}[(C\epsilon)(C\epsilon)'] = C\mathbb{E}[\epsilon\epsilon']C' = C\Omega C' = C(C'C)^{-1}C' = CC^{-1}C'^{-1}C' = I$$

and the following estimator is BLUE

$$\begin{aligned}\hat{\beta}_{\mathrm{GLS}} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \\ &= (X'C'CX)^{-1}X'C'Cy \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y\end{aligned}$$

Variance of GLS

$$
\begin{aligned}
\mathbb{V}[\hat{\beta}_{\text{GLS}}] &= \mathbb{V}[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\beta+\epsilon)] \\
&= \mathbb{V}[\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon] \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\mathbb{V}[\epsilon]\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\
&= (X'\Omega^{-1}X)^{-1}
\end{aligned}
$$

Adjusting cross-section estimations for GLS

- Given the variance of residuals

$$\Omega = \frac{1}{T}\left(\beta\Sigma_f\beta' + \Sigma\right)$$

$$\hat{\lambda}_{\text{GLS}} = (\beta'(\beta\Sigma_f\beta' + \Sigma)^{-1}\beta)^{-1}\beta'(\beta\Sigma_f\beta' + \Sigma)^{-1}\mathbb{E}_T[R_i^e]$$

Trick to simplify the formula

Sherman–Morrison–Woodbury formula

$$(A+UCV)^{-1}=A^{-1}-A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$$
$$(\Sigma+\beta\Sigma_f\beta')^{-1}=\Sigma^{-1}(I-\beta(\beta'\Sigma^{-1}\beta+\Sigma_f^{-1})^{-1}\beta'\Sigma^{-1})$$

And define idempotent matrix M as

$$\begin{aligned}
M &= I-\beta(\beta'\Omega^{-1}\beta)^{-1}\beta'\Omega^{-1} \\
M^2 &= (I-\beta(\beta'\Omega^{-1}\beta)^{-1}\beta'\Omega^{-1})(I-\beta(\beta'\Omega^{-1}\beta)^{-1}\beta'\Omega^{-1}) \\
&= I-2\beta(\beta'\Omega^{-1}\beta)^{-1}\beta'\Omega^{-1}+\beta(\beta'\Omega^{-1}\beta)^{-1}\beta'\Omega^{-1}\beta(\beta'\Omega^{-1}\beta)^{-1}\beta'\Omega^{-1} \\
M\beta &= \beta-\beta(\beta'\Omega^{-1}\beta)^{-1}\beta'\Omega^{-1}\beta=0
\end{aligned}$$

Continuation.

$$
\begin{aligned}
(\Sigma + \beta\Sigma_f\beta')^{-1} &= (\Sigma + \beta\Sigma_f\beta')^{-1}M^{-1}M \\
&= \Sigma^{-1}(I - \beta(\beta'\Sigma^{-1}\beta + \Sigma_f^{-1})^{-1}\beta'\Sigma^{-1})M^{-1}M \\
&= \Sigma^{-1}(M - M\beta(\beta'\Sigma^{-1}\beta + \Sigma_f^{-1})^{-1}\beta'\Sigma^{-1})M \\
&= \Sigma^{-1}M^{-1}M \\
&= \Sigma^{-1}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\hat{\lambda}_{\text{GLS}} &= (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\mathbb{E}_T[R_i^e] \\
\mathbb{V}[\hat{\lambda}_{\text{GLS}}] &= \frac{1}{T}\left(\beta'(\beta\Sigma_f\beta' + \Sigma)^{-1}\beta\right)^{-1} \text{ does not drop term } \Sigma_f
\end{aligned}
$$

The cancelation of Σ_f in some of the formulas only occurs if you have terms in the denominator and numerator.

Continuation

$$\mathbb{V}[\hat{\lambda}_{\mathrm{GLS}}] = \frac{1}{T}\left(\beta'(\beta\Sigma_f\beta' + \Sigma)^{-1}\beta\right)^{-1}$$

$\hat{\alpha}$ variance under GLS

$$
\begin{aligned}
\hat{\alpha}_{\text{GLS}} &= \mathbb{E}_T[R^e] - \beta\hat{\lambda}_{\text{GLS}} \\
&= \mathbb{E}_T[R^e] - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\mathbb{E}_T[R_i^e] \\
&= (I - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1})\mathbb{E}_T[R_i^e] \\
\mathbb{V}[\hat{\alpha}_{\text{GLS}}] &= \frac{1}{T}\left(I - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\right)(\beta\Sigma_f\beta' + \Sigma)\left(I - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\right)' \\
&= M(\beta\Sigma_f\beta' + \Sigma)M \\
&= M\Sigma M \\
&= \frac{1}{T}\left(I - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\right)\Sigma\left(I - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\right)'
\end{aligned}
$$

Im sure there is a simplified version of this formula (try to find it).