

1 Factor Return Regression

1.1 Barra Regression Formulation

- **Excess return** \mathbf{er}_t , where $\mathbf{er}_t = \mathbf{r}_t - rf_t$, \mathbf{r}_t is the stock return and rf_t is the risk-free return; the default choice of which should be the 3-month treasury yield.
- **Factor Exposures** $\mathbf{x}_{s,t}$ the exposures to style factor s , $\mathbf{x}_{i,t}$ the exposures to industry factor i
- **Specific Return** $\epsilon_t = \mathbf{er}_t - \sum_s \mathbf{x}_{s,t-1} f_{s,t} - \sum_i \mathbf{x}_{i,t-1} f_{i,t}$, where $f_{s,t}$ is the factor return for style s and $f_{i,t}$ is the factor return for style i that are to be estimated at time t .
- **Regression Weights** $\mathbf{w}_{j,t}$ is the corresponding regression weights for ticker j at time t .

The target of Barra Regression is to obtain the estimates for factor returns via a weighted regression with certain constraints. The objective function of the regression model is to minimize the weighted sum of squares of the specific returns, which is:

$$\min_{f_{s,t}, f_{i,t}} \sum_j w_{j,t} \epsilon_{j,t}^2$$

The corresponding constraints are applied to the industry factor returns only, where they are assumed to be cancelled out:

$$\sum_i f_{i,t} = 0$$

Denote \mathbf{W}_t as a diagonal matrix for regression weights such that $\mathbf{W}_t = \text{diag}(w_{t,j})$, $\mathbf{X}_{S,t} = [\mathbf{x}_{s,t} \forall s \in S]$ as the loading matrix for style factor exposures and $\mathbf{X}_{I,t} = [\mathbf{x}_{i,t} \forall i \in I]$ as the loading matrix for industry factor exposures, we can re-write the regression problem in matrix form as:

$$\begin{aligned} \min_{F_{s,t}, F_{i,t}} \quad & \frac{1}{2} (\epsilon_t - \mathbf{X}_{S,t} F_{S,t} - \mathbf{X}_{I,t} F_{I,t})^\tau \mathbf{W} (\epsilon_t - \mathbf{X}_{S,t} F_{S,t} - \mathbf{X}_{I,t} F_{I,t}) \\ & \mathbf{1}^\tau F_{I,t} = 0 \end{aligned}$$

The Barra Regression Model is essentially a **Quadratic Programming with Linear Constraints** problem, which have an explicit solution when the constraints are equality constraints.

1.2 Explicit Solutions

For a quadratic programming problem with linear constraints:

$$\begin{aligned} \min_x & \frac{1}{2}x^\tau Qx + p^\tau x \\ \text{s.t.} & A^\tau x = q \end{aligned}$$

If $Q \succ 0$, namely, Q is positive-definite and non-singular, we shall have an explicit solution such that:

$$\hat{x} = -Q^{-1}p - Q^{-1}A(A^\tau QA)^{-1}A^\tau Q^{-1}p + Q^{-1}A(A^\tau QA)^{-1}q$$

If Q is singular, in order to obtain \hat{x} , we need to solve the following linear system corresponding to the Lagrangian dual problem, which is:

$$\begin{bmatrix} Q & A \\ A^\tau & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -p \\ q \end{bmatrix}$$

If the block matrix on the left hand side in the linear system above is still non-singular, then the optimal solution to this quadratic programming problem would not be unique.

2 Barra Factor Covariance Estimation

After obtaining the factor returns F_t at time t , we need to estimate the corresponding factor return covariance matrix Σ_t , which is actually constructed using the long-run covariance estimation. Assuming that we have window size τ , where only $F_t, F_{t-1}, \dots, F_{t-\tau}$ are used for obtaining the covariance, a maximal lag of $l(\tau)$ for including auto-correlations and that all factor returns F_t are centered with mean value removed:

$$\begin{aligned} \tau \Sigma_t &= \sum_{t-\tau \leq i, j \leq t} \mathcal{K}\left(\frac{i-j}{l(\tau)}\right) F_i F_j^\tau \\ &= \mathcal{K}(0) \sum_{t-\tau \leq i, j \leq t} F_i F_i^\tau + \sum_{k=1}^{l(\tau)} \mathcal{K}\left(\frac{k}{l(\tau)}\right) \left(\sum_{t-\tau \leq i-k < i \leq t} (F_i F_{i-k}^\tau + F_{i-k} F_i^\tau) \right) \end{aligned}$$

The function $\mathcal{K}(\cdot)$ is a kernel function which should satisfy that $\mathcal{K}(0) = 1$, $\mathcal{K}(u) = 0$ $|u| \geq 1$, and $\mathcal{K}(\cdot)$ is even, continuous and differentiable almost everywhere. This is essentially using kernel density estimator to obtain the spectral density of time series at zero-frequency in the frequency domain. Moreover, for the maximal auto-correlation lag, $l(\tau)$, it should be relatively small compared to τ , $l(\tau) = o(\tau)$.

In the Barra Model, the kernel function is chosen as the Bartlett-kernel function, which is equivalent to the one used in the Newey-West heteroskedastic covariance estimator in regression models. Specifically speaking, the Bartlett kernel function is:

$$\mathcal{K}(u) = 1 - |u|$$

3 Specific Risk

3.1 Specific Risk Estimation

After obtaining the specific returns ϵ_t , we can instantly have the raw time-series based specific risk estimation σ_t^{TS} , where for ticker j $\sigma_{j,t}^{TS} = std(\epsilon_{j,t})$. In order to remove the outliers, we regress the raw estimators onto the factor exposures loading matrix:

$$\ln(\sigma_t^{TS}) \sim \mathbf{X}_{t-1}\beta_t$$

The structured specific risk σ_t^{STR} is then obtain as:

$$\sigma_t^{STR} = E_0 \cdot \exp(\mathbf{X}_{t-1}\beta_t)$$

We shall also have the blended specific risk estimator for ticker j as:

$$\hat{\sigma}_{j,t} = \gamma_j \sigma_{j,t}^{TS} + (1 - \gamma_j) \sigma_{j,t}^{STR}$$

For most tickers, γ_j is set with a value of 1 to retain as much information as possible. Only the tickers with abnormal specific risks (outliers) will be given a value γ_j smaller than one, which serves the purpose of smoothing the values toward the structured specific risks.

3.2 Specific Risk Shrinkage

The Bayesian shrinkage specific risk is the blended specific risk shrunk towards the market-cap weighted mean:

$$\sigma_{j,t}^{SH} = v_{j,t} \bar{\sigma}_t + (1 - v_{j,t}) \hat{\sigma}_{j,t}$$

where $\bar{\sigma}_t$ is the corresponding market cap weighted mean of blended specific risk:

$$\bar{\sigma}_t = \sum_j m_{j,t} \hat{\sigma}_{j,t}$$

From where, the corresponding shrinkage intensity $v_{j,t}$ can be constructed as:

$$v_{j,t} = \frac{q|\hat{\sigma}_{j,t} - \bar{\sigma}_t|}{\Delta_t + q|\hat{\sigma}_{j,t} - \bar{\sigma}_t|}$$

$$\Delta_t = \sqrt{\frac{\sum_j (\hat{\sigma}_{j,t} - \bar{\sigma}_t)^2}{n}}$$

In the actual implementation, q is the parameter that requires careful tuning to control the level of shrinkage intensity, which decides how much weight should be put on the mean of specific risk for each individual ones.