# 1 The Model

## 1.1 Generalized Least Squares (GLS) Regression

Consider the model of data

$$\begin{aligned} y &= X\beta + \epsilon \\ \epsilon &\sim \mathcal{N}(\epsilon|0, \Sigma) \end{aligned} \tag{1}$$

Assuming $X$ is given, we have $y|X \sim \mathcal{N}(y|X, \beta, \Sigma)$ and conditional log-likelihood is:

$$l_{y|X}(\beta) = \log f(y|X;\beta) \propto -\frac{1}{2}(y - X\beta)^T \Sigma^{-1}(y - X\beta)$$

To find $\hat{\beta}^{MLE} = \text{argmax}_\beta \, l_{y|X}(\beta)$ we solve

$$\nabla_\beta l_{y|X}(\beta) = -X^T\Sigma^{-1}X\beta + X^T\Sigma^{-1}y = 0$$

Hence

$$\hat{\beta}^{MLE} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y \tag{2}$$

# 2 Implementation for multi-variate return time series

Consider the problem

$$\begin{aligned} R_{t,i} &= \sum_{k=1}^{K} \beta_k A_{t,i}^{(k)} + \epsilon_{t,i}, \\ (t, i) &\in \{(1, i_{1,1}), ..., (1, i_{1,N_1}), \\ &\quad \cdots \\ &\quad (T, i_{T,1}), ..., (T, i_{T,N_T})\} \end{aligned} \tag{3}$$

In the above:

- $R_{t,i}$ are returns for stock $i$ over period $t$;
- $A_{t,i}^{(k)}$ are features, usually alphas or their lagged versions.

Note that indices are 2-dimensional, but not necessarily rectangular, i.e. set of stocks can be different on different days. For notation convenience we introduce daily (random) vectors where we put all variables with same index $t$ in a column[1]:

$$R_{t,\cdot} = \begin{bmatrix} R_{t,i_{t,1}} \\ \vdots \\ R_{t,i_{t,N_t}} \end{bmatrix}, A_{t,\cdot}^{(k)} = \begin{bmatrix} A_{t,i_{t,1}}^{(k)} \\ \vdots \\ A_{t,i_{t,N_t}}^{(k)} \end{bmatrix}, \epsilon_{t,\cdot} = \begin{bmatrix} \epsilon_{t,i_{t,1}} \\ \vdots \\ \epsilon_{t,i_{t,N_t}} \end{bmatrix} \tag{4}$$

and we can further stack daily vectors:

[1] note that dimensions for vectors with different index $t$ might be different

$$R = \begin{bmatrix} R_{1,\cdot} \\ \vdots \\ R_{T,\cdot} \end{bmatrix}, A^{(k)} = \begin{bmatrix} A^{(k)}_{1,\cdot} \\ \vdots \\ A^{(k)}_{T,\cdot} \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_{1,\cdot} \\ \vdots \\ \epsilon_{T,\cdot} \end{bmatrix} \tag{5}$$

so we can rewrite Equation 3 in vector form:

$$R = A\beta + \epsilon, \epsilon \sim \mathcal{N}(\epsilon|0, \Sigma) \tag{6}$$

$$A = \left[A^{(1)} \cdots A^{(k)}\right] \tag{7}$$

## 3 Exploting special structure of errors $\epsilon$

In practice, we can assume special structure of $\Sigma$ that let us simplify the implementation:

- errors for different time periods are uncorrelated $cov(\epsilon_{t_1,i}, \epsilon_{t_2,j}) = 0$ if $t_1 \neq t_2$.

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_T \end{bmatrix} \tag{8}$$

- errors in the same period $t$ have covariance matrix $\Sigma_t$ that has a "low-rank plus diagonal" structure

$$\Sigma_t = L_t \Sigma_t^f L_t^T + \Lambda_t. \tag{9}$$

We can exploit these properties to simplify Equation 2:

- inverse of $\Sigma$ is

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_1^{-1} & 0 & \cdots & 0 \\ 0 & \Sigma_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_T^{-1} \end{bmatrix} \tag{10}$$

- each $\Sigma_t$ can be inverted using Woodbury inversion formula:

$$\Sigma_t^{-1} = (L_t \Sigma_t^f L_t^T + \Lambda_t)^{-1} = \Lambda_t^{-1} + \Lambda_t^{-1} L_t((\Sigma_t^f)^{-1} + L^T \Lambda_t^{-1} L) L_t^T \Lambda_t^{-1} \tag{11}$$

Thus computation Equation 2 reduces to:

$$\beta^{GLS} = (\sum_{t=1}^{T})^{-1} \tag{12}$$