1 Generalized Least Squares Regression

1.1 Basic model

Consider the model of data

$$
\begin{aligned}
&y = X\beta + \epsilon \qquad (1)\\
&\epsilon \sim \mathcal{N}(\epsilon|0, \Sigma)\\
&X \in \mathbb{R}^{n\times k}, y \in \mathbb{R}^n, \epsilon \in \mathbb{R}^n
\end{aligned}
$$

In the above, $\Sigma \in \mathbb{S}^n_{++}$ is a known covariance matrix of errors, $\beta \in \mathbb{R}^k$ is an unknown constant vector of coefficients. Our goal is to find a good estimate of β.

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

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

Definition 1. $\hat{\beta}^{GLS}$ is MLE estimate of β in the Model 1:

$$
\hat{\beta}^{GLS} = \underset{\beta}{\operatorname{argmax}}\, l(\beta) \qquad (2)
$$

Definition 1 gives an optimization problem

$$
\underset{\beta}{\text{maximize}} -\frac{1}{2}(y - X\beta)^T\Sigma^{-1}(y - X\beta) \qquad (3)
$$

the first-order condition $\nabla_\beta l(\beta) = 0$ gives *normalized equations*:

$$
X^T\Sigma^{-1}X\beta = X^T\Sigma^{-1}y
$$

Solution to the Problem 6 can be written as

$$
\hat{\beta}^{GLS} = (X^T\Sigma^{-1}X)^\dagger X^T\Sigma^{-1}y \qquad (4)
$$

1.2 Comparison to OLS

Looking at Equation 6 one can note that GLS is equivalent to OLS if data $D = (X, y)$ is preprocessed. For example, using Cholesky decomposition $\Sigma = LL^T$ we can lef-tmultiply the equation by L^{-1}:

$$
\begin{aligned}
&y' = X'\beta + \epsilon' \qquad (5)\\
&y' = L^{-1}y, X' = L^{-1}X\\
&\epsilon' = L^{-1}\epsilon \sim N(\epsilon'|0, I)
\end{aligned}
$$

and the Problem 6 reduces to the usual OLS problem:

$$
\underset{\beta}{\text{maximize}} -\frac{1}{2}(y' - X\beta)^T(y' - X'\beta) \Leftrightarrow \underset{\beta}{\text{minimize}}\, \frac{1}{2}\left\|y' - X'\beta\right\|_2^2 \qquad (6)
$$

1.3 Extensions

One can extend the basic model by considering coefficient vector β as random and introducing prior distribution[1].

If $f(\beta) \propto \mathit{const}$ we get same model as above

[1]Equivalently, introduce regularization terms directly into Problem 6.

2 Implementation for multi-variate return time series

Consider the problem

$$
\begin{aligned}
R_{t,i} = & \sum_{k=1}^{K} \beta_k A_{t,i}^{(k)} + \epsilon_{t,i}, \\
(t,i) \in & \{(1, i_{1,1}), ..., (1, i_{1,N_1}), \\
& \dots \\
& (T, i_{T,1}), ..., (T, i_{T,N_T})\}
\end{aligned}
\tag{7}
$$

In the above:

- $R_{t,i}$ are returns for stock i over period t;
- $A_{t,i}^{(k)}$ are features, for example, alpha vectors, their lags, subuniverses, etc.

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

$$
R_{t,\cdot} = \begin{bmatrix} R_{t,i_{t,1}} \\ \vdots \\ R_{t,i_{t,N_t}} \end{bmatrix}, A_{t,\cdot}^{(k)} = \begin{bmatrix} A_{t,i_{t,1}}^{(k)} \\ \vdots \\ A_{t,i_{t,N_t}}^{(k)} \end{bmatrix}, \epsilon_{t,\cdot} = \begin{bmatrix} \epsilon_{t,i_{t,1}} \\ \vdots \\ \epsilon_{t,i_{t,N_t}} \end{bmatrix} \tag{8}
$$

and we can further stack daily vectors:

$$
R = \begin{bmatrix} R_{1,\cdot} \\ \vdots \\ R_{T,\cdot} \end{bmatrix}, A^{(k)} = \begin{bmatrix} A_{1,\cdot}^{(k)} \\ \vdots \\ A_{T,\cdot}^{(k)} \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_{1,\cdot} \\ \vdots \\ \epsilon_{T,\cdot} \end{bmatrix} \tag{9}
$$

so we can rewrite Equation 7 in vector form:

$$
R = A\beta + \epsilon, \epsilon \sim \mathcal{N}(\epsilon|0, \Sigma) \tag{10}
$$

$$
A = \left[A^{(1)} \cdots A^{(k)}\right] \tag{11}
$$

3 Exploting special structure of Σ

In practice, we can assume special structure of Σ that allows simplified implementation of Equation 14.

3.1 No autocorrelation of errors

It is reasonable to assume that errors for different time periods don't correlate $cov(\epsilon_{t_1,i}, \epsilon_{t_2,j}) = 0$ if $t_1 \neq t_2$, thus.

$$
\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_T \end{bmatrix} \tag{12}
$$

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

Inversion of block diagonal matrix is simple:

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_1^{-1} & 0 & \cdots & 0 \\ 0 & \Sigma_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_T^{-1} \end{bmatrix} \tag{13}$$

Plugging Equation 9 into Equation 14 and using Equation 13 translates into:

$$\hat{\beta}^{GLS} = \sum_{t=1}^{T} (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y \tag{14}$$

3.2 Factor structure of errors in the same period

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

$$\Sigma_t = L_t \Sigma_t^f L_t^T + \Lambda_t. \tag{15}$$

each Σ_t can be inverted using Woodbury inversion formula:

$$\Sigma_t^{-1} = (L_t \Sigma_t^f L_t^T + \Lambda_t)^{-1} = \Lambda_t^{-1} + \Lambda_t^{-1} L_t ((\Sigma_t^f)^{-1} + L^T \Lambda_t^{-1} L) L_t^T \Lambda_t^{-1} \tag{16}$$

Thus computation Equation 14 reduces to:

$$\beta^{GLS} = \left(\sum_{t=1}^{T} A_{t,\cdot}{}^T \Sigma_t^{-1} A_{t,\cdot} \right)^{-1} \tag{17}$$