1. Different rate have different length
2. The simple LOESS method do not consider volatility clustering
3. May use PCA to detect outlier
4. But one crucial fact: these are time series, so they are correlated serially.

Difficulties in **MTS (multivariate time series)** analysis

* Too many parameters with high dimension
* Identifiability problem: hard to estimate the precise value of parameters

Possible solutions:

* Vector auto-regression model (VAR)
* Kronecker index and Scalar components- by using Canonical correlation analysis and likelihood ratio test
* Factor models- by using PCA, LASSO, K-means clustering etc.
* Mahalanobis distance: simple, but not consider the property of time-dependency

1. Combine all the time series
2. Plot the series together
3. Perform PCA- dimension reduction? Is it meaningful?

**LOESS: local regression** (locally weighted polynomial regression)

Idea:

* combine simplicity of linear regression + flexibility of nonlinear regression, fit segments of the data
* Any complex function can be well approximated by a small neighborhood by lower-order polynomials

Procedure:

1. For each data point, fit a low-degree polynomial regression to some subset of the data.
2. Polynomial is fitted using weighted least square: more weights to the near points.
3. Parameter:
   1. Smoothing parameter: proportion of data used in each fit, smaller, capture more local behavior.
   2. Degree of local polynomial: 1 or 2
   3. Weight function: nearer, more weight (more likely to be related)

Advantage of LOESS:

* No need to assume function to fit
* Flexible

Disadvantages of LOESS:

* Not produce a function that can be represented by a math formula
* Prone to outlier: how to avoid

My thoughts on LOESS:

* Only on Univariate
* After fit line, use what criteria to detect outlier? Just look at boxplot outlier is bad, since a cluster of outlier may not be outlier

Use **PCA (Principal Component Analysis)** to detect outlier

1. Advantage of dimension reduction
2. But lose time dependency