Multivariate Pairs Trading with Structural Change Detections in Cointegrated Relationships

K.-L. WANG[1], M.-C. Shiue[2], and T.-S. Dai[3]

National Chiao Tung University

Master Thesis Defense

[1]E-mail: polyphonicared@gmail.com
[2]Advisor
[3]Co-Advisor

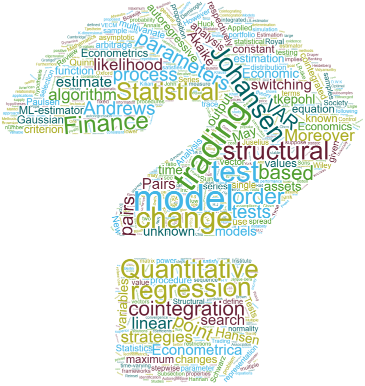
Sshort Biodata

Kuan-Lun Wang, a master student, majors in the generalized pairs trading. The major goal of his research is to develop a algorithmic trading mechanism based on statistical arbitrage. His expertise include automatic search procedures for model selection, multivariate co-integration approach, and structural change test.

Research Interests

His research interests are time series models, simulation modeling, and portfolio choice. A central theme of application is the study of multivariate pairs trading in real time, searching assets with long-run equilibrium, and building riskless portfolios. Much current work involves structural change analysis and co-integration test of finite order vector autoregressive process and estimates the probability of mean reversion. Such methods are important in a variety of applications, including economic indicators and hedging. One specific example is the index funds tied to indexes with more lower costs and more lower risks.

Outline



Should you have any questions, feel free to contact us.

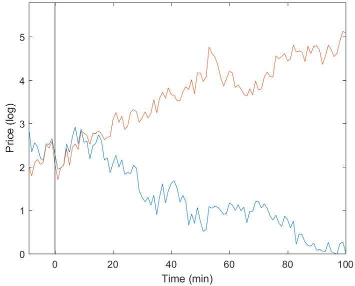
Background (1/2)

1. Pairs trading strategies (PTSs)

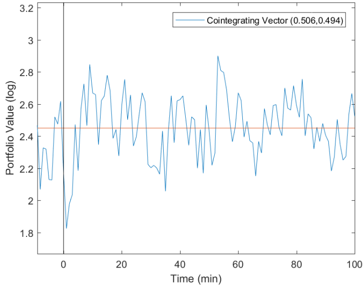
> The goal of a PTS is to earn a greater profit without suffering risks in probability.

2.
3. The statistical arbitrage goal,
4. Two-step process:
   1. Find two assets with prices that move together, historically;
   2. Monitor the subsequent trading spread between the assets. Should the spread diverge, short the higher priced asset and long the lower priced asset until the spread coverges again.
5. Multivariate frameworks [52]:
   - Portfolios of assets against other portfolios of assets.

Background (2/2)



(a) Log Prices Process



(b) Mean Reversion Process

Figure: An Example for PTSs

Approaches of PTSs

1. Distance, cointegration, time series, stochastic control, machine learning (ML), combined forecasts, copula, principal components analysis (PCA), etc.
   - Others: distance approach [27], time-series approach [24], etc.
   - Intent: equilibrium relationships [82, 24, 27, 48].
     - E.g., $\beta' y_t = \beta_1 y_{1t} + \cdots + \beta_K y_{Kt} + c = 0$.
2. Vector autoregressive (VAR) model:

   Where u_t is a white noise process with time-varying covariance matrix $\Sigma_u(t)$.

   $y_t = v + \sum_{i=1}^{p} A_i y_{t-i} + u_t$,
   - Vector error correction model (VECM) representation [25]:

     $\Delta y_t = \alpha\beta' y^*_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t$.
3. Tests and estimator for stable linear combinations,
   - Reduced rank regression (RRR) techniques [7],
     - RRR: $z_{0t} = \alpha\beta' z_{1t} + \Gamma z_{2t} + u_t$.
   - Extensions on likelihood-based theory [39, 43].

Problem (1/2)

1. L denotes the lag operator, ω denotes the magnitudes of the disturbance, m denotes the location of the outlier, and

$$\xi_t^m = \begin{cases} 1, & \text{if } t = m, \\ 0, & \text{otherwise.} \end{cases}$$

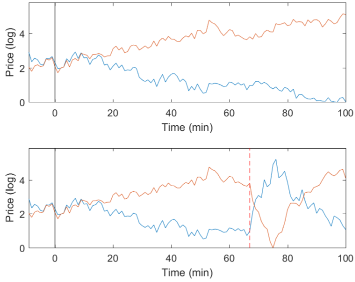
   - Level changes (LC): $y_t = X_t + \frac{1}{1-L}\omega\xi_t^m$.

- Switching regression models [8]:
   - Two-Regime: $y_t = \begin{cases} X_t\beta_1 + u_t, & \text{if } t > t_0, \\ X_t\beta_2 + u_t, & \text{if } t \leq t_0. \end{cases}$

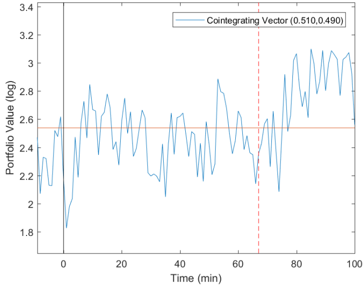
2. Switching low order VAR model:

$$y_t = \begin{cases} v_1 + \sum_{i=1}^{p} A_{i1} y_{t-i} + u_t, & \text{if } t > t_0, \\ v_2 + \sum_{i=1}^{p} A_{i2} y_{t-i} + u_t, & \text{if } t \leq t_0. \end{cases}$$

Problem (2/2)



(a) Two-Regime Model



(b) False Estimator

Figure: An Example of PTSs with Single Structural Change Point

Study First Aim

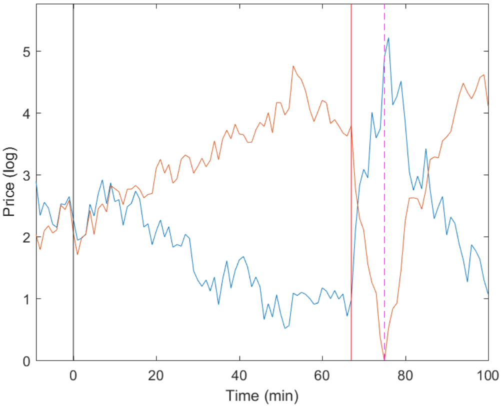


Figure: Test and Estimator for Structural Change

Measur

Let $p_{i,t}$ and $p_{j,t}$ be denoted realizations of the normalized price processes $P_i = (P_{i,t})_{t \in T}$ and $P_j = (P_{j,t})_{t \in T}$ of the securities i and j of a pair and $s^2(\cdot)$ sample variance. Empirical spread variance $s^2_{P_i - P_j}$ can be expressed as

$$s^2_{P_i - P_j} = \frac{1}{T} \sum_{t=1}^{T} (p_{i,t} - p_{j,t})^2 \left(\frac{1}{T} \sum_{t=1}^{T} (p_{i,t} - p_{j,t}) \right).$$

We can solve for the average sum of squared distances $\overline{ssd}_{P_i,P_j}$ in the formation period:

$$\overline{ssd}_{P_i,P_j} = \frac{1}{T} \sum_{t=1}^{T} (p_{it} - p_{jt})^2 = s^2_{P_i - P_j} + \left(\frac{1}{T} \sum_{t=1}^{T} (p_{i,t} - p_{j,t}) \right)^2.$$

- $\overline{ssd}_{P_i,P_j} = T^{-1} \sum_{t=1}^{T} (p_{it} - p_{jt})^2$ (standardizing).
- Multivariate frameworks:
  - One of the problems is the significant computational power required to assess potential pairs of portfolios, owing to the enormous pool of assets;
  - The second goal of this study is to propose a searching cointegrated price search procedure in the VAR model by taking advantage of the forward and backward stepwise regression procedures [63, 69, 22, 53].

Static Setting (1/6)

Model: A K-Dimensional Switching VAR(p) Model

The logarithmic asset price processes are formulated as a K-dimensional switching VAR(p) model, with time-varying parameters given as follows:

$$y_t = \nu(t) + \sum_{i=1}^{p} A_i(t) y_{t-i} + u_t, \quad t = 1, \ldots, T,$$

where u_t is a white noise process with time-varying covariance matrix $\Sigma_u(t)$.

Static Setting (2/6)

Model: A K-Dimensional Switching VAR(p) Model –Continued

The parameters are given by

$$\begin{aligned} \nu(t) &= \nu_1 1_1(t) + \nu_2 1_2(t) \\ A_i(t) &= A_{i1} 1_1(t) + A_{i2} 1_2(t), \quad i = 1, \ldots, p, \\ \Sigma_u(t) &= \Sigma_{u1} 1_1(t) + \Sigma_{u2} 1_2(t), \end{aligned}$$

where $\pi_0 T$ is, if any, the unknown structural change point, $1_1(t)$ and $1_2(t)$ are indicator functions of subsample $[1, \pi_0 T]$ and $(\pi_0 T, T]$, respectively; ν_1 and ν_2 are $(K \times 1)$ intercepts; A_{i1} and A_{i2} are the $(K \times K)$ constant parameters for $i = 1, \ldots, p$.

Static Setting (3/6)

Model: VECM Representation

Moreover, we consider that the model has a VECM representation

$$\Delta y_t = \alpha(t)\beta(t)' y^*_{t-1} + \sum_{i=1}^{p-1} \Gamma_i(t)\Delta y_{t-i} + u_t, \quad t = 1, \dots, T,$$

where

$$\begin{aligned} y^*_{t-1} &= (y'_{t-1}, 1)', \\ \alpha(t)\beta(t)' &= \alpha_1\beta_1' 1_1(t) + \alpha_2\beta_2' 1_2(t), \\ \Gamma_i(t) &= \Gamma_{i1} 1_1(t) + \Gamma_{i2} 1_2(t), \quad i = 1, \dots, p-1, \end{aligned}$$

where $\alpha_j\beta_j' = (-(I_K - \sum_{i=1}^{p} A_{ij}), \nu_j)$ for $j = 1, 2$,
$\Gamma_{ij} = -\sum_{k=1}^{p-i} A_{i+k,j}$ for $i = 1, \dots, p-1$ and $j = 1, 2$.

Static Setting (4/6)

Model: VECM Representation–Continued

Then $\alpha(t)$ and $\beta(t)$ are the $(K \times r)$ adjustment coefficients and the $((K+1) \times r)$ cointegration parameters, respectively; and the process y_t denotes cointegration of rank r. The VECM model has no $\alpha(t)\beta(t)'$ or $\Gamma_i(t)$ if $r = 0$ or $p \leq 1$, respectively.

Proposition: Well-Defined

The switching VAR(p) model can be derived from the VECM representation, and vice versa.

Static Setting (5/6)

Proof.

The process defining y_t can be written as $y_t = \Delta y_t + y_{t-1}$. Then VECM model, with Δy_t replaced by $y_t - y_{t-1}$, implies that

$$
\begin{aligned}
y_t =& \nu(t) + \sum_{i=1}^{p} A_i(t) y_{t-1} - \sum_{i=1}^{p-1} \sum_{k=1}^{p-i} A_{i+k}(t) y_{t-i} \\
&+ \sum_{i=1}^{p-1} \sum_{k=1}^{p-i} A_{i+k}(t) y_{t-i-1} + u_t.
\end{aligned}
$$

Static Setting (6/6)

Proof–Continued.

Introducing $j = i + 1$, we get

$$y_t = \nu(t) + \sum_{i=1}^{p} A_i(t) y_{t-1} - \sum_{i=1}^{p-1} \sum_{k=1}^{p-i} A_{i+k}(t) y_{t-i} + \sum_{j=2}^{p} \sum_{k=1}^{p-j+1} A_{j+k-1}(t) y_{t-i} + u_t.$$

Combining the coefficients of y_{t-i}, $i = 1, \ldots, p$ we obtain the switching VAR(p) model.

Common Assumptions (1/6)

Assumption 1: Sample

The time series $y_1, \dots, y_T$ is available; that is, we have a sample of size T.

Assumption 2: Initial Values

Presamples are also available; that is, we have initial values.

Assumption 3: Stability or Instability

The process y_t is stable or integrated of order 1.

What are the stable process and the integrated process?

Common Assumptions (2/6)

Definition: Convergent Linear Process [43]

A linear process is defined by $y_t = \sum_{i=0}^{\infty} C_i u_{t-i}$, $t = 0, 1, \ldots,$ where $C(z) = \sum_{i=0}^{\infty} C_i z^i$ is convergent for $|z| \leq 1 + \delta$ for some $\delta > 0$.

Definition: Stable Process [43]

A stochastic process y_t which satisfies that $y_t - E[y_t] = \sum_{i=0}^{\infty} C_i u_{t-i}$ is stable if $C = \sum_{i=0}^{\infty} C_i \neq 0$.

Definition: Integrated Process [43]

A st

$d =$

The degenerate case, where there is no random element, e.g., $X(t) = mt$, where m is a constant, is also a stable process. (Wiki: Stable Process)

However, we need more assumptions to give "nice" process.

Commo

Johansen (1991) considered a VAR model with Gaussian errors written in the error correction form

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-i} + \Phi D_t + \mu + \epsilon_t$$

where D_t are seasonal dummies orthogonal to the constant term. Further, ...

Assu

The
$|C_2(z$
$C_1(z) = I_K - \sum_{i=1}^{p} A_{i1} z^i$ and $C_2(z) = I_K - \sum_{i=1}^{p} A_{i2} z^i$.

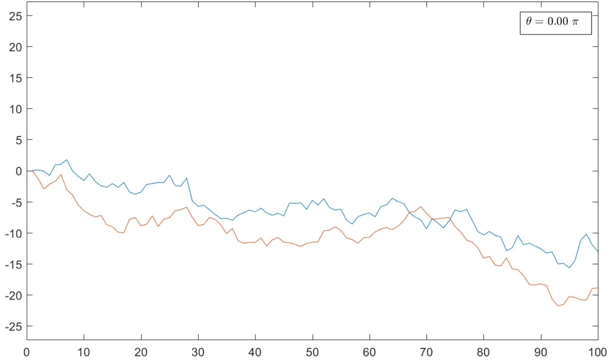
In other words, we consider the switching VAR(p) model that exclude explosive roots and seasonal roots other than $z = 1$.

Proposition: Granger's Representation Theorem [39]

Under assumption 4, the processes $\beta_1' y_t$ and $\beta_2' y_t$ is stationary.

What are the seasonal roots and the stationary process?

Common Assumptions (4/6)



$$y_t = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} y_{t-1} + \epsilon_t \text{ where } \epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1) \text{ and } y_1 = 0$$

Figure: Seasonal Root Examples

Common Assumptions (5/6)

Definition: (Strictly) Stationary Process [43]

By a stationary process, we mean a process for which the distribution of $y_{t_1}, \ldots, y_{t_m}$ is the same as the distribution of $y_{t_1+h}, \ldots, y_{t_m+h}$ for any $h = 1, 2, \ldots$.

Definition: Cointegrating Rank [43]

Let y_t be integrted of order 1. We call y_t cointegrated with cointegrating vector $\beta \neq 0$ if $\beta' y_t$ can be made stationary by a suitable choice of its initial distribution. The cointegrating rank is the number of linearly independent cointegrating relations, and the space spanned by the cointegrating relations is the cointegrating space.

Common Assumptions (6/6)

To ensure the test for detecting structural change is workable, we follow Andrews1993's (1993) suggestion of assuming that the structural change point is not too close to the samples' beginning (time 1) or ending (time T). Thus, both subsamples $[1, \pi_0 T]$ and $(\pi_0 T, T]$ contain sufficient data for the exponential likelihood ratio (LR) test.

Assu

If the

interval $T\Pi$.

Assumption 6: Stability of Parameters

No structural change if $\alpha(t)$, $\beta(t)$, and $\Gamma_i(t)$ are stable at time $T\Pi$.

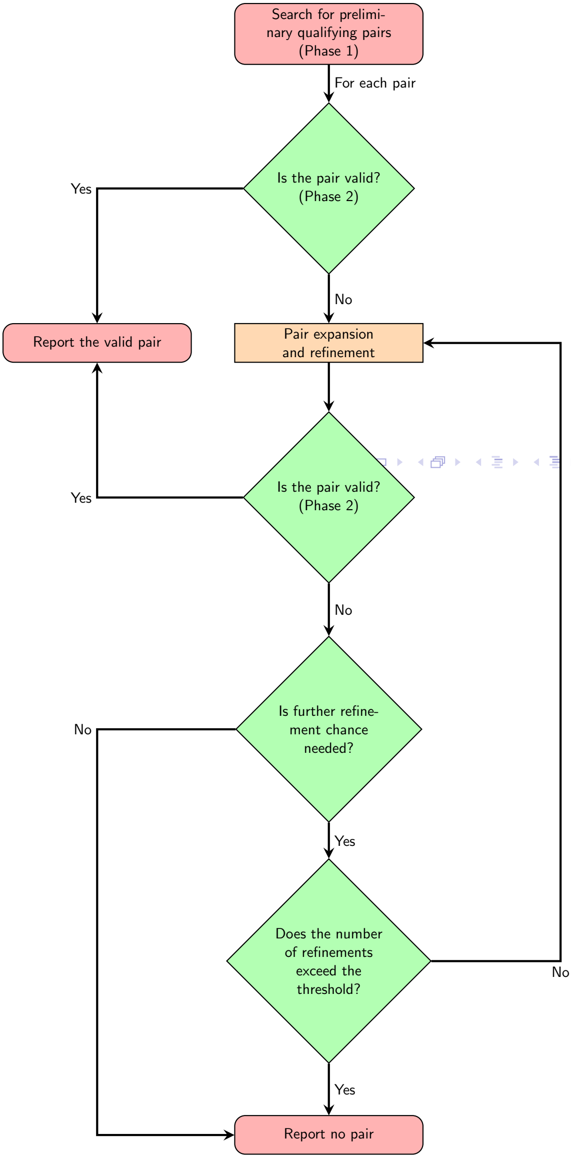
How can we estimate parameters in the model?

Measure Terms and Overview

1. Measure terms:
    - Information criterion value,
    - Cintegration test p-value,
        - Null hypothesis is no cointegration vector.
    - Unknown single structural change test (USSC test) p-value,
        - Null hypothesis is no change point.
    - VAR normality test p-value.
        - Null hypothesis is normality.
2. A oversimplifying our cointegrated pairs search algorithm:
    1. Identifying qualifying pairs;
    2. Pair validation;
    3. Pair expansion and refinement.

Overview of our Cointegrated Pairs Search Algorithm

Figure: Overview of our Cointegrated Pairs Search Algorithm



Identifying Qualifying Pairs (1/5)

Give

1. An enormous pool,
2. A predetermined number, N,
3. An exogenous probability s,
4. Two-stage Bernoulli sampling process.

Proposition: Converge to Bernoulli Distribution

Let the random variable z be the number of trials required to observe the N-th successful trial under Bernoulli trials, with success probability p; that is, z has a negative binomial distribution. Let z_t equal the number of trials required to observe the N-th success by uniformly sampling t objects without replacement. Then, the sequence z_t converges in distribution to z with an increment of t.

Identifying Qualifying Pairs (2/5)

Proof.

Suppose z has the cumulative distribution function

$$F_z(x) = \sum_{i=N}^{x} \binom{i-1}{N-1} p^N (1-p)^{i-N}, \quad x = N, N+1, N+2, \ldots,$$

and z_t has the cumulative distribution function

$$F_{z_t}(x) = \sum_{i=N}^{x} \binom{x-1}{N-1} \left(\prod_{j=1}^{i} \frac{\lfloor pt \rfloor - j}{t} \right) \left(\prod_{j=i+1}^{x} \left(1 - \frac{\lfloor pt \rfloor - j}{t} \right) \right)$$

for $t \geq N$.

Identifying Qualifying Pairs (3/5)

Proof–Continued.

Then,

$$\begin{aligned}
&\lim_{t\to\infty} F_{z_t}(x) \\
\leq &\lim_{t\to\infty} \sum_{i=N}^{x} \binom{x-1}{N-1} \left(\prod_{j=1}^{i} \frac{\lfloor pt \rfloor}{t} \right) \left(\prod_{j=i+1}^{x} \left(1 - \frac{\lfloor pt \rfloor - x}{t} \right) \right) \\
= &F_z(x)
\end{aligned}$$

and the result

$$\lim_{t\to\infty} F_{z_t}(x) = F_z(x)$$

for all $x = N, N+1, N+2, \ldots$. Therefore, $F_{z_t} \to F_z$ as $t \to \infty$, and z_t converges in distribution to z.

Identifying Qualifying Pairs (4/5)

First Stage in Two-Stage Bernoulli Sampling Process

1. Picks $N_{\text{criterion}}$ pairs with an information criterion for model fitting ranked as the top $p_{\text{criterion}}$ percentage from all possible trading pairs,
2. Uniformly samples $x_{\text{criterion}}$ pairs from all pairs and selecting the top $N_{\text{criterion}}$ pairs with better information criterion values from sampled pairs.

The proposition suggests the minimum number of $x_{\text{criterion}}$ in order to achieve the goal of the first stage with an exogenous probability s.

Identifying Qualifying Pairs (5/5)

Second Stage in Two-Stage Bernoulli Sampling Process

1. We picks $N_{\text{portfolios}}$ pairs cointegration test p-values ranked in the top p_{CI} percent of the pairs generated in the first stage,
2. The proposition again suggests the minimum number of samples x_{CI} to be drawn uniformly from $N_{\text{criterion}}$ pairs generated in the first stage in order to achieve the second-stage goal with probability s.
3. Since the null hypothesis of the cointegration test is the absence of a cointegration relation.

the cointegration property.

PTSs Validation

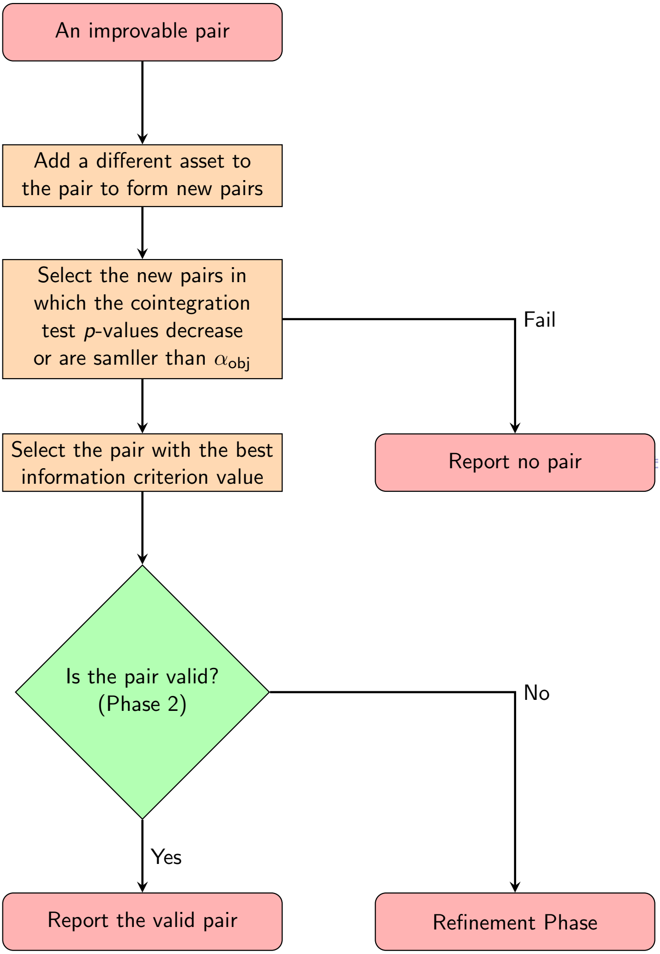
Definition: PTSs Validation

Give an objective test p-value α_{obj}. Then a pairs is valid if

1. the pair cointegration test p-value small than α_{obj},
2. the pair USSC test p-value not small than α_{obj},
3. the pair normality test p-value not small than α_{obj}.

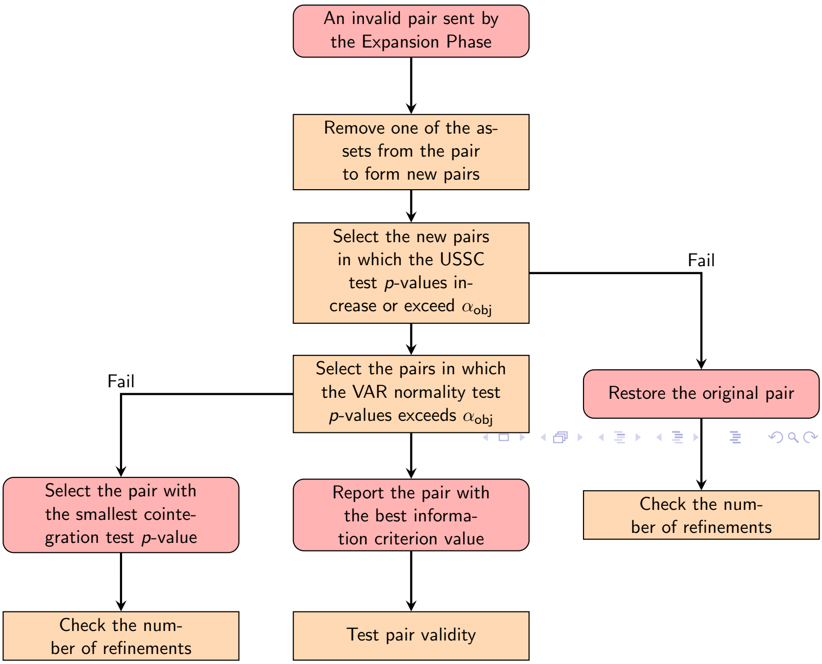
Overview of pair Expansion and Refinement Phase (1/2)

Figure: Overview of the Expansion Phase



Overview of PTSs Expansion and Refinement Phase (2/2)

Figure: Overview of the Refinement Phase



Order Estimators and Information Criterions

1. Criteria for VAR order selection:
   - Likelihood ratio (LR) testing procedures [67],
   - Final prediction error criterion (FPE) [3, 4],
   - Akaike's information criterion (AIC) [5, 6],
   - Schwarz criterion (SC) [74],
   - Hannan-Quinn criterion (HQ) [30].
2. Consistency of VAR(p) model order estimator [30, 71, 66],
   - X: FPE and AIC,
   - O: SC and HQ.
3. The best,
   - Low order VAR: SC $>$ LR,FPE,AIC,HQ [57].

What is the SC?

VAR(p) Model

Definition: VAR(p) [58]

Let y_t be a K-dimensional process as in

$$y_t = \nu + \sum_{i=1}^{p} A_i y_{t-i} + u_t, \quad t = 0, \pm 1, \pm 2, \ldots,$$

where ν is a $(K \times 1)$ constant vector, $A_1, \ldots, A_p$ are $(K \times K)$ constant matries with $A_p \neq 0$, and u_t is a white noise process with the covariance matrix Σ_u. Then the process y_t is a VAR(p) process.

Estimators for VAR Model

Proposition: The LS and Gaussian ML Estimators of Σ_u [58]

The LS estimator and the Gaussian ML estimator of Σ_u is

$$\hat{\Sigma}_u = \frac{1}{T} \ldots (y_t - Bx_t)',$$

The LS and Gauss … ors of $(\nu, A_1, \ldots, A_p)$ are the B.

$$x_t = \begin{pmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}$$

where $B = \left(\sum_{t=1}^{T} y_t x_t'\right) \ldots^{-1}$ and $x_t = (1, y'_{t-1}, \ldots, y'_{t-p})'$.

Schwarz Criterion (SC)

Definition: Schwarz Criterion (SC) [74]

Let p be an unknown positive integer, and let M be a known constant which $p \leq M$. Thus, the SC is defined as

$$\text{SC}(m) = \log \left|\hat{\Sigma}_u(m)\right| + m\frac{K^2 \log T}{T},$$

where $\hat{\Sigma}_u(m)$ is the Gaussian ML estimator of Σ_u by fitting a VAR(m) model for $m = 0, \ldots, M$, and the estimate $\hat{p}$ of p based on the SC is

$$\hat{p} = \arg\min\{\text{SC}(m) : m = 0, \ldots, M\}.$$

Consistency of VAR Order Estimators

Proposition: Consistency of VAR Order Estimators [58]

Let y_t be a K-dimensional stationary, stable VAR(p) process with standard white noise (that is, u_t is independent white noise with bounded fourth moments). Suppose the maximum order $M \geq p$ and $\hat{p}$ is chosen so as to minimize a criterion

$$\text{Cr}(m) = \log |\hat{\Sigma}_u(m)| + m\frac{c_T}{T}$$

over $m = 0, 1, \ldots, M$. Here $\hat{\Sigma}_u(m)$ denotes the (quasi

$K^2 \log T$

estimator of Σ_u obtained for a VAR(m) model and c_T is a nondecreasing sequence of real numbe

$$\frac{K^2 \log T}{T} \to 0$$

on the sample size T. Then $\hat{p}$ is consistent if

$K^2 \log T \to \infty$

$$c_T \to \infty \text{ and } \frac{}{T} \to 0 \text{ as } T \to \infty.$$

Measur

The power of a binary hypothesis test is the probability that the test correctly rejects the null hypothesis (H0) when the alternative hypothesis (H1) is true. It can be equivalently thought of as the probability of accepting the alternative hypothesis (H1) when it is truethat is, the ability of a test to detect an effect, if the effect actually exists. That is,

$$\text{power} = \Pr(\text{reject } H_0 | H_1 \text{ is true}).$$

(Wiki: Statistical power)

1.
   - Small sample power [59],
     - trace test [39, 43] > maximum eigenvalue test [45, 43].
   - Problem,
     - These tests are based on Gaussian ML estimator.
2. For normality test in VAR model,
   - Jarque-Bera normality test [50].

What are the trace test and the Jarque-Bera normality test?

Vector Error Co

Table 1. Models and LR-type tests.

Determ. terms	Model	Test statistic	Reference
$\mu_0 = \mu_1 = 0$	$\Delta y_t = \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	LR^0	Johansen (1988, 1995)
μ_0 arbitrary	$\Delta y_t = \nu_0 + \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	LR^{i0}	Johansen (1991)
$\mu_1 = 0$	$\Delta y_t = \Pi^* \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	LR^*	Johansen and Juselius (1990)
	$\Delta y_t = \Pi(y_{t-1} - \mu_0) + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	LR^{SL}	Saikkonen and Luukkonen (1997) Saikkonen and Lütkepohl (2000a)
μ_0, μ_1 arbitrary	$\Delta y_t = \nu + \Pi^+ \begin{bmatrix} t-1 \\ y_{t-1} \end{bmatrix} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	LR^+	Johansen (1992, 1994, 1995)
	$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	LR^{PC}	Perron and Campbell (1993)
	$\Delta y_t - \mu_1 = \Pi(y_{t-1} - \mu_0 - \mu_1(t-1))$ $+ \sum_{j=1}^{p-1}$	LR^{GLS}	Saikkonen and Lütkepohl (2000a) (2000)

Definition: VE

The VAR(p) pr

$$\Delta y_t = \alpha\beta' y_{t-1}^* + \sum_{i=1}$$

$$\alpha\beta' = \begin{pmatrix} -(I_K - \sum^{p} A_i) \\ \nu \end{pmatrix}$$

$$y_{t-1}^* = \begin{pmatrix} y_{t-1} \\ 1 \end{pmatrix}$$

where $\alpha\beta' = (-(I_K - \sum_{i=1}^{p} A_i)', \nu')'$, $y_{t-1}^* = (y_{t-1}', 1)'$, and $\Gamma_i = -\sum_{j=1}^{p-i} A_{i+j}$ for all $i = 1, \ldots, p-1$, and the loading matrix of VECM and the cointegration matrix are α and β, respsectively.

Reduced Rank Regression

Definition: Reduced Rank Re

$$\begin{pmatrix} \Delta y_{t-1} \end{pmatrix} \quad \begin{pmatrix} y_{t-1} \end{pmatrix}$$

$$\begin{array}{cccccccccc} & \Delta y_t & = & \alpha\beta' & \begin{pmatrix} y_{t-1} \\ 1 \end{pmatrix} & + & (\Gamma_1, \dots, \Gamma_{p-1}) & \begin{pmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix} & + & u_t \\ \Longrightarrow & z_{0t} & = & \alpha\beta' & z_{1t} & + & \Gamma & z_{2t} & + & u_t \end{array}$$

$$z_{0t} = \alpha\beta' z_{1t} + \Gamma z_{2t} + u_t, \qquad t = 1, \dots, T.$$

Kronecker Product (1/2)

Definition: Kronecker Product [58]

Let $A = (a_{ij})$ and $B = (b_{ij})$ be $(m \times m)$ and $(p \times q)$ matrices, respectively. The $(mp \times nq)$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B. \end{pmatrix}$$

is the Kronecker product (or direct product) of A and B.

Kronecker Product (2/2)

Example: An Example for Kronecker Product [58]

The Kronecker product of $A = \begin{pmatrix} 3 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 5 & -1 \\ 3 & 3 \end{pmatrix}$ is

$$A \otimes B = \begin{pmatrix} 15 & -3 & 20 & -4 & -5 & 1 \\ 9 & 9 & 12 & 12 & -3 & -3 \\ 10 & -2 & 0 & 0 & 0 & 0 \\ 6 & 6 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$B \otimes A = \begin{pmatrix} 15 & 20 & -5 & -3 & -4 & 1 \\ 10 & 0 & 0 & -2 & 0 & 0 \\ 9 & 12 & -3 & 9 & 12 & -3 \\ 6 & 0 & 0 & 6 & 0 & 0 \end{pmatrix}.$$

vec Operator (1/2)

Definition: vec Operator [58]

Let $A = (a_1, \ldots, a_n)$ be an $(m \times n)$ matrix with $(m \times 1)$ columns a_i. The *vec operator* transforms A into an $(mn \times 1)$ vector by stacking the columns, that is,

$$\text{vec}(A) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

vec Operator (2/2)

Example: An Example for vec Operator [58]

For instance, if $A = \begin{pmatrix} 3 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 5 & -1 \\ 3 & 3 \end{pmatrix}$, then

$$\text{vec}(A) = \begin{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} \\ \begin{pmatrix} 4 \\ 0 \end{pmatrix} \\ \begin{pmatrix} -1 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 4 \\ 0 \\ -1 \\ 0 \end{pmatrix} \text{ and } \text{vec}(B) = \begin{pmatrix} \begin{pmatrix} 5 \\ 3 \end{pmatrix} \\ \begin{pmatrix} -1 \\ 3 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ -1 \\ 3 \end{pmatrix}.$$

Gaussi

Prop

Furt

wher
defin
the c
respe
the C

$\hat{\beta} =$

$\hat{\alpha} =$

respectively.

The log likelihood function is given apart from a constant by

$$\log L = \log L(\phi, \alpha, \beta) = -\frac{1}{2} T \log |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{T} (z_{0t} - \alpha\beta' z_{1t} - \Gamma z_{2t})' \times \Sigma_u^{-1} (z_{0t} - \alpha\beta' z_{1t} - \Gamma z_{2t})$$

The first order conditions for estimating Γ are given by

$$\sum_{t=1}^{T} (z_{0t} - \alpha\beta' z_{1t} - \Gamma z_{2t}) z_{2t}' = 0$$

We write the conditions as

$$M_{02} = \alpha\beta' M_{12} + \Gamma M_{22}$$

such that

$$\Gamma = \Gamma(\alpha, \beta') = M_{02} M_{22}^{-1} - \alpha\beta' M_{12} M_{22}^{-1}.$$

This leads to the definition of the residuals

$$R_{0t} = z_{0t} - M_{02} M_{22}^{-1} z_{2t},$$

$$R_{1t} = z_{1t} - M_{12} M_{22}^{-1} z_{2t}.$$

Asympt

Johansen and Juselius (1990) showed the maximal eigenvalue statistic

$$\lambda_{max} \xrightarrow{d} \text{maxeig}\left(\int_0^1 (\mathrm{d}\, B_t) F_t' \left(\int_0^1 F_t F_t' \,\mathrm{d}\, t\right)^{-1} \int_0^1 F_t (\mathrm{d}\, B_t)'\right)$$

where $\lambda_{max} = T\lambda_{r+1}$, and maxeig$(M)$ denotes the marginal eigenvalue of the matrix M, and the LR-like test can be conducted for the pair of hypotheses

$$H_0\colon \text{rk}(\alpha\beta') = r \text{ against } H_1\colon \text{rk}(\alpha\beta') = r + 1.$$

Prop

The

where $\lambda_{trace} = \sum_{i=r+1}^{K} \lambda_i$, B_t is a $(K - r)$-dimensional Brownian motion, and $F_t = (B_t', 1)'$.

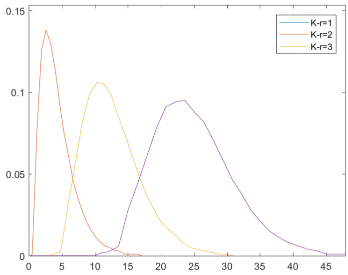
Johansen and Juselius's (1990) simulations are performed replacing the Brownian motion by a Gaussian random walk with 400 steps and the process is stimulated 6,000 times.

Simu

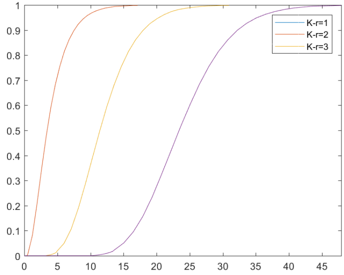
We g

on 100,000 simulation repetitions and 3,600 steps.

Asymptotic Distribution of Trace Test Statistic (2/2)



(a) PDF



(b) CDF

Figure: The Asymptotic Distribution of Trace Statistics

Test for Cointegration

Test for Cointegration Rank [39, 43]

$$H_0\colon \mathrm{rk}(\alpha\beta') = r_0 \text{ against } H_1\colon r_0 < \mathrm{rk}(\alpha\beta') \leq K.$$

Test Strategy of cointegration test [58]

Test a sequence of null hypotheses,

$$H_0\colon \mathrm{rk}(\alpha\beta') = 0, \ldots, H_0\colon \mathrm{rk}(\alpha\beta') = K - 1.$$

However, these estimator and tests for cointegration based on the Gaussian property.

How can we test for normality?

Normality

Definition: Normality [33]

The random variable X has a normal distribution of its p.d.f is defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty$$

where μ and σ are parameters satisfying $-\infty < \mu < \infty$ and $0 < \sigma < \infty$.

Definition: K-Variate Normality [29]

Let X be a $(K \times 1)$ random vector whose elements X_i are independent identically $N(0,1)$ distributed. Then the distribution of X is called K-variate (standard) normal distribution.

Measures of Distribution (1/2)

Definition: Skewness Measure [33]

One measure of the asymmetry of the distribution is skewness and is defined by

$$\frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}} = \frac{E[(X-\mu)^3]}{\sigma^3}.$$

When a distribution is symmetrical about the mean, the skewness is equal to zero.

Proposition: Normality has Zero Skewness Measure [33]

The normal distribution is asymmetry that immediately that the skewness of it is 0.

Measures of Distribution (2/2)

Definition: Kurtosis Measure [58]

One measure of the tailedness of the distribution is Kurtosis and is defined by

$$\frac{E[(X-\mu)^4]}{E[(X-\mu)^2]^{4/2}} - 3 = \frac{E[(X-\mu)^4]}{\sigma^4} - 3.$$

The reason to subtract off 3 is that the normal distribution has zero Kurtosis measure.

Normality Proposition

Proposition: An Asymptotic Distribution about Normality [50]

Let $X_1, \ldots, X_N$ be are independent and have K-variate normal distributions. Then

$$\sqrt{T}\begin{pmatrix} b_1 \\ b_2 - 3_K \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} 6I_K & 0 \\ 0 & 24I_K \end{pmatrix}\right)$$

where $b_1 = (b_{11}, \ldots, b_{1K})'$ and $b_2 = (b_{21}, \ldots, b_{2K})'$ with $b_{1i} = T^{-1}\sum_{n=1}^{N} X_{1n}^3$ and $b_{2i} = T^{-1}\sum_{n=1}^{N} X_{2n}^4$ for $i = 1, \ldots, K$.

Chi-Square (χ^2) Distribution and Tests for Normality

$$
\begin{array}{rccc}
 & \sqrt{T}\begin{pmatrix} b_1 \\ b_2 - 3_K \end{pmatrix} & \xrightarrow{d} & N\left(0, \begin{pmatrix} 6I_K & 0 \\ 0 & 24I_K \end{pmatrix}\right) \\
\Longrightarrow & \sqrt{T}b_2 - 3_K & \xrightarrow{d} & N(0, 24I_K) \\
\Longrightarrow & \frac{\sqrt{T}}{\sqrt{24}}(b_2 - 3_K) & \xrightarrow{d} & N(0, I_K) \\
\Longrightarrow & \left(\frac{\sqrt{T}}{\sqrt{24}}(b_2 - 3_K)\right)'\left(\frac{\sqrt{T}}{\sqrt{24}}(b_2 - 3_K)\right) & \xrightarrow{d} & \chi^2_K \\
\Longrightarrow & K = T(b_2 - 3_K)'(b_2 - 3_K)/24 & \xrightarrow{d} & \chi^2_K
\end{array}
$$

Take $S = Tb_1'b_1/6$ and $K = T(b_2 - 3_K)'(b_2 - 3_K)/24$. Then the asymptotic distributions of S ans K are

$$S \xrightarrow{d} \chi^2_K \text{ and } K \xrightarrow{d} \chi^2_K.$$

Moreover, define the Jarque-Bera normality test statistic $J = S + K$, and the asymptotic distribution of J is

$$J \xrightarrow{d} \chi^2_{2K}.$$

Jarque-Bera normality test for VAR(p)

Proposition: LS Estim… …vels VAR(p) [58]

The LS and (… …stimators of $(A_1, \dots, A_p)$ are the B.

$$x_t = \begin{pmatrix} y_{t-1} \\ \dots \\ y_{t-p} \end{pmatrix}$$

The LS (…

$\tilde{u}_t = y_t - Bx_t$ where … $)(\sum_{t=1}^{T} x_t x_t')^{-1}$ and $x_t = (y'_{t-1}, \dots, y'_{t-p})'$.

Proposition: Jarque-Bera Normality Test for VAR(p) [50]

Define $\tilde{w}_t = (\tilde{w}_{1t}, \dots, \tilde{w}_{Kt})' = \tilde{P}^{-1}\tilde{u}_t$ where $\tilde{P}$ is the lower triangular matrix with positive diagonal satisfying $\tilde{P}\tilde{P}' = \tilde{\Sigma}_u$ where $\tilde{\Sigma}_u = (T - Kp - 1)^{-1} \sum_{t=1}^{T} (\tilde{u}_t - \bar{u}_t)(\tilde{u}_t - \bar{u}_t)'$ with sample mean $\bar{u}$ of $\tilde{u}_t$. Then,

$$\tilde{J} \equiv \tilde{S} + \tilde{K} \xrightarrow{d} \chi^2_{2K}$$

where $\tilde{S} = T\tilde{b}_1'\tilde{b}_1/6$ and $\tilde{K} = T(\tilde{b}_2 - 3_K)'(\tilde{b}_2 - 3_K)/24$ where $\tilde{b}_1 = (\tilde{b}_{11}, \dots, \tilde{b}_{1K})'$ and $\tilde{b}_2 = (\tilde{b}_{21}, \dots, \tilde{b}_{2K})'$ with $\tilde{b}_{1i} = T^{-1} \sum_{t=1}^{T} \tilde{w}_{it}^3$ and $\tilde{b}_{2i} = T^{-1} \sum_{t=1}^{T} \tilde{w}_{it}^4$ for $i = 1, \dots, K$.

Measure of a Single Structural Change

1. Estimators of the switching cointegrated VAR [31],
   - Two convergence properties:
     - Convergence rate,
     - Converge to the global or local optimum.
   - Switching algorithms,
     - Boswijk (1995) [15] > Johansen and Juselius (1992) [46]
2. Test for structural changes in cointegrated VAR model:
   - With a known single structural changes [31],
     - Based on Boswijk (1995) [15].
   - With an unknown single structural change,
     - Power [8, 10]:
       Exp LR [10] > Avg LR [10] > Sup LR [8] > CUMSUM [17],
     - Based on LR test statistic.

How can we test USSC in cointegrated VAR model?

Another Reduced Rank Regression

$$Z_{1t} = \begin{pmatrix} z_{1t}1_1(t) \\ z_{1t}1_2(t) \end{pmatrix}$$

$$Z_{2t} = \begin{pmatrix} z_{2t}1_1(t) \\ z_{2t}1_2(t) \end{pmatrix}$$

Definition: Another Reduced Ran

Define $Z_{1t} = (z'_{1t}1_1(t), z'_{1t}1_2(t))'$ and $\Gamma = (\Gamma_1, \Gamma_2)$ wh
$\Gamma_i = (\Gamma_{1i}, \ldots, \Gamma_{p-1,i})$ for $i = 1, 2$, and $Z_{2t} = (z_{2t}1_1(t)', z_{2t}1_2(t)')'$.

$$\begin{aligned} z_{0t} &= (\alpha_1, \alpha_2)\begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} z_{1t}1_1(t) \\ z_{1t}1_2(t) \end{pmatrix} + (\Gamma_1, \Gamma_2) \begin{pmatrix} z_{2t}1_1(t) \\ z_{2t}1_2(t) \end{pmatrix} + u_t \\ \Longrightarrow \quad z_{0t} &= \alpha\beta' \; Z_{1t} + \Gamma \; Z_{2t} + u_t \end{aligned}$$

$$z_{0t} = \alpha\beta' Z_{1t} + \Gamma Z_{2t} + u_t, \qquad t = 1, \ldots, T$$

where $\alpha = (\alpha_1, \alpha_2)$, $\beta = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}$.

Notation: Meaning of RRR

The above RRR shows the following regression switching RRR:

$$z_{0t} = \alpha_1\beta'_1 z_{1t}1_1 + \alpha_2\beta'_2 z_{1t}1_2 + \Gamma_1 z_{2t}1_1 + \Gamma_2 z_{2t}1_2 + u_t$$

Restriction in Switching cointegrated VAR Model (1/2)

Notation: Observed vec Operator of Parameters

The vec operator implies

$$\text{vec}(\alpha, \Gamma) = \begin{pmatrix} \text{vec}(\alpha) \\ \text{vec}(\Gamma) \end{pmatrix} = \begin{pmatrix} \text{vec}(\alpha_1) \\ \text{vec}(\alpha_2) \\ \text{vec}(\Gamma_1) \\ \text{vec}(\Gamma_2) \end{pmatrix}$$

and

$$\text{vec}(\beta) = \begin{pmatrix} \text{vec}(\beta_1) \\ \text{vec}(0_{(K+1)\times r}) \\ \text{vec}(0_{(K+1)\times r}) \\ \text{vec}(\beta_2) \end{pmatrix} = \begin{pmatrix} \text{vec}(\beta_1) \\ 0_{(2Kr+2r)\times 1} \\ \text{vec}(\beta_2) \end{pmatrix}.$$

Restriction in Switching cointegra

$$\left(\mathrm{vec}(\alpha_1)\right.$$

v

$$\mathrm{vec}(\beta) = \begin{pmatrix} \mathrm{vec}(\beta_1) \\ 0_{(2Kr+2r)\times 1} \\ \mathrm{vec}(\beta_2) \end{pmatrix} = H\phi$$

Restriction: Stable Parameters versus

The restriction forms: $\mathrm{vec}(\alpha, \Gamma) = G\psi$ and $\mathrm{vec}(\beta) = H\phi$.

- With a single structural change:

$$G_0 = I_{2(Kr+K^2(p-1))} \text{ and } H_0 = \begin{pmatrix} I_{(K+1)r} & 0_{(K+1)r} \\ \multicolumn{2}{c}{0_{2(K+1)r}} \\ 0_{(K+1)r} & I_{(K+1)r} \end{pmatrix}.$$

- Without structural change:

$$G_1 = \begin{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes I_{Kr} & 0_{2Kr\times K^2(p-1)} \\ 0_{2K^2(p-1)\times Kr} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes I_{K^2(p-1)} \end{pmatrix} \text{ and } H_1 = \begin{pmatrix} I_{(K+1)r} \\ 0_{2(K+1)r\times(K+1)r} \\ I_{(K+1)r} \end{pmatrix}.$$

Estimator of the Switching Cointegrated VAR Model (1/3)

Proposition: Gaussian ML Estimator of Parameters [31]

The ML estimator of α, β, and Γ satisfy the restrictions $\text{vec}(\alpha, \Gamma) = G\psi$ and $\text{vec}(\beta) = H\phi$ satisfy

$$\begin{aligned}\text{vec}(\hat{\alpha}, \hat{\Gamma}) =& G\left(G' \sum_{t=1}^{T}\left(\begin{pmatrix}\hat{\beta}' Z_{1t} Z_{1t}' \hat{\beta} & \hat{\beta}' Z_{1t} Z_{2t}' \\ Z_{2t} Z_{1t}' \hat{\beta} & Z_{2t} Z_{2t}'\end{pmatrix} \otimes \hat{\Sigma}_{ut}^{-1}\right) G\right)^{-1} \\ & \times G' \sum_{t=1}^{T} \text{vec}(\hat{\Sigma}_{ut}^{-1} z_{0t} (Z_{1t}' \hat{\beta}, Z_{2t}')),\end{aligned}$$

and

$$\begin{aligned}\text{vec}(\hat{\beta}) =& H\left(H' \sum_{t=1}^{T}\left(\hat{\alpha}' \hat{\Sigma}_{ut}^{-1} \hat{\alpha} \otimes Z_{1t} Z_{1t}'\right) H\right)^{-1} \\ & \times H' \sum_{t=1}^{T} \text{vec}(Z_{1t} (z_{0t} - \hat{\Gamma} Z_{2t})' \hat{\Sigma}_{ut}^{-1} \hat{\alpha}).\end{aligned}$$

Estimator of the Switching Cointegrated VAR Model (2/3)

Proposition: Gaussian ML Estimator of Parameters [31] –Continued

The Gaussian ML estimator of Σ_{ut} satisfy the restrictions $\text{vec}(\alpha, \Gamma) = G\psi$ and $\text{vec}(\beta) = H\phi$ satisfy

$$\hat{\Sigma}_{ui} = (\Delta T_i)^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' 1_i, \quad i = 1, 2$$

where $\Delta T_1 = \pi_0 T$, $\Delta T_2 = T - \pi_0 T$, $\hat{u}_t = z_{0t} - \hat{\alpha}\hat{\beta}' Z_{1t} - \hat{\Gamma} Z_{2t}$, and the maximum value of the likelihood function is given by

$$L_{\max}(\hat{\alpha}, \hat{\beta}, \hat{\Gamma}, \hat{\Sigma}_{ut}) = (2\pi)^{-Tp/2} \prod_{t=1}^{T} |\Sigma_{ut}|^{-1/2} \exp\left(-\frac{Tp}{2}\right).$$

Estimator of the Switching Cointegrated VAR Model (3/3)

Algorithm: Gaussian ML estimated Iterative Algorithm [31]

1. Give the initial value of $\{\hat{\alpha} = \hat{\alpha}_0, \hat{\beta} = \hat{\beta}_0, \hat{\Gamma} = (\hat{\Gamma}_0, \hat{\Gamma}_0), \hat{\Sigma}_{u1} = \hat{\Sigma}_{u0}, \hat{\Sigma}_{u2} = \hat{\Sigma}_{u0}\}$.
2. For fixed values of $\hat{\beta}$, $\hat{\Sigma}_{u1}$, and $\hat{\Sigma}_{u2}$ estimate $\hat{\alpha}$ and $\hat{\Gamma}$.
3. For fixed values of $\hat{\alpha}$, $\hat{\Gamma}$, $\hat{\Sigma}_{u1}$, and $\hat{\Sigma}_{u2}$ estimate $\hat{\beta}$.
4. For fixed values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\Gamma}$ estimate $\hat{\Sigma}_{u1}$ and $\hat{\Sigma}_{u2}$.
5. Repeat the steps until the value of the maximum likelihood

In th
likeli
500 iterations has been reached.

In software program CATS, the algorithm stops when $|\Delta L_{max,i}| \leq 10^{-6}$ or after a maximum of 500 iterations has been reached. When needed, these default values can be change by the user.

Test for a Known Single Structural Change

Definition: LR Test Statistic [58]

Let M_1 be a submodel of model M_0, and the values of maximum likelihood function for M_0 and M_1 are $L_{max,0}$ and $L_{max,1}$. Then the LR test statistic λ_{LR} is $-2\log L_{max,1}/L_{max,0}$.

Proposition: Asymptotic Distribution of LR Tests [31]

If M_1 is a submodel of model M_0 with q fewer parameters, then χ^2_q is the asymptotic distribution of the LR test of M_1, tested against M_0.

Notation: Degrees of Freedom for χ^2 Distribution

The degrees of freedom for χ^2 in our case is $f_G + f_H$ where $f_G = \text{rk}(G_0) - \text{rk}(G_1)$ and $f_H = \text{rk}(H_0) - \text{rk}(H_1)$.

Test for an Unknown Single Structural Change (1/2)

Proposition: Asymptotic Distribution of LR Statistics [10]

For one model in the time interval $[1, T]$ with an unknown single structural change point in $T\Pi \subsetneq [1, T]$, we define the statistic $\lambda_{LR,\pi}$ is the LR test statistic for $\pi \in \Pi$. Then

$$\lambda_{Exp} \equiv \log \Big($$

$$\begin{aligned} \lambda_{Sup} &= \max_{\pi\in\Pi} \lambda_{LR,\pi} &\xrightarrow{d}& \max_{\pi\in\Pi} Q(\pi) \\ \lambda_{Avg} &= \frac{1}{\#(T\Pi)} \sum_{\pi\in\Pi} \lambda_{LR,\pi} &\xrightarrow{d}& \frac{1}{|\Pi|} \int_{\pi\in\Pi} Q(\pi)\, d\pi \end{aligned}$$

where λ_{Exp} is the exponential LR (Exp LR) test statistic and $Q(\pi) = (\tilde{B}_\pi - \pi\tilde{B}_1)'(\tilde{B}_\pi - \pi\tilde{B}_1)/(\pi(1-\pi))$ where $\tilde{B}_\pi$ is the q-dimensional Brownian motion.

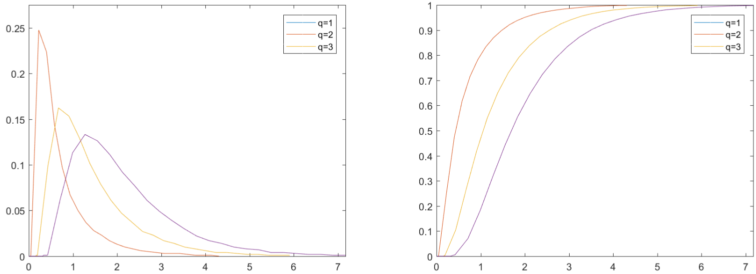
Andrews (2003) gave critical values for values of p in $\{1, \ldots, 20\}$ based on 100,000 simulation repetitions and 3,600 values of gird of steps for sup LR.

Nota

We g

repetitions and 3,600 values of grid of $[0, 1]$.

Test for an Unknown Single Structural Change (2/2)



When no knowledge of the change point is available, Andrews (1993) suggested using a restricted interval Π, such as $\Pi = [0.15, 0.85]$ for sup LR test.

Figure: The Asymptotic Distribution of Exp-LR Statistics for $\Pi = [0.15, 0.85]$

Trading Strategy

Definition: USSC strategy

Give

- a trading time bound h.
- a specifying arbitrage probability $q_{\text{arbitrage}}$,
- an estimated arbitrage probability q_{est},
- a transaction cost at the chosen level of b_{cost} basis points per trade per dollar,

The USSC strategy:

- if $q_{\text{est}} \geq q_{\text{arbitrage}}$, then we can build a portfolio in which the return is irrelevant to the market return with a sufficiently high probability by investing both assets with weights defined by the cointegration vector component as the investment proportions;
- if $q_{\text{est}} < q_{\text{arbitrage}}$, then we do not build any portfolio.

Arbitrage Probability Estimator (1/5)

Review: Logarithm USSC Pair Pric

Suppose a logarithm USSC pair pri $((K+1)\times r)$ cointegration parameters $\beta_0 = (\beta_{01}$ ar combination asset process $(\beta'_{0i}y^*_t)'$ is normal and stationary for $i = 1, \ldots, r$.

$$\beta'_0 y^*_t = \begin{pmatrix} \beta'_{01}y^*_t \\ \vdots \\ \beta'_{0r}y^*_t \end{pmatrix}$$

1. There is no loss of generality
2. Suppose two random vectors
   - $Y_1 = (\beta'_{0r_0}y_1, \ldots, \beta'_{0r_0}y_T)'$,
   - $Y_2 = (Y_{21}, \ldots, Y_{2h})' = (\beta'_{0r_0}y_{T+1}, \ldots, \beta'_{0r_0}y_{T+h})'$.
   - $Y \sim N(\mu, \Sigma)$ where

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},\ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},\ \text{and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$Y_1 = \begin{pmatrix} \beta'_{0r_0}y_1 \\ \vdots \\ \beta'_{0r_0} \end{pmatrix}$$

$$Y_2 = \begin{pmatrix} Y_{21} \\ \vdots \\ Y_{2h} \end{pmatrix} = \begin{pmatrix} \beta'_{0r_0}y_{T+1} \\ \vdots \\ \beta'_{0r_0}y_{T+h} \end{pmatrix}$$

Arbitrage Probability Estimator (2/5)

Proposition: Marginal and Conditional Distributions of a Multivariate Normal [73, 47, 58, 32]

Let y_1 and y_2 be two random vectors such that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

where the partitioning of the mean vector and covariance matrix corresponds to the vector $(y_1', y_2')'$. Then,

1. $y_2 \sim N(\mu_2, \Sigma_{22})$,
2. the conditional distribution of y_2 given $y_1 = c$ is

$$(y_2 | y_1 = c) \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(c - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

Arbitrage Probability Estimator (3/5)

Proposition: Marginal and Conditional Distributions of a Multivariate Normal [73, 47, 58, 32] –Continued

Let y_1 and y_2 be two random vectors such that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

where the partitioning of the mean vector and covariance matrix corresponds to the vector $(y_1', y_2')'$. Then,

3. If Σ_{11} is singular, the inverse can be replaced by a generalized inverse;
4. y_1 and y_2 are independent if and only if $\Sigma_{12} = \Sigma_{21}' = 0$.

Arbitrage Probability Estimator (4/5)

1. The conditional distribution of Y_2 given $Y_1 = c$, where c is the real observed price vector is multivariate normal,

$$(Y_2 | Y_1 = c) \sim N(\mu_{2|1}, \Sigma_{2|1}),$$

where

$$\mu_{2|1} = \begin{pmatrix} \mu_{21|1} \\ \vdots \\ \mu_{2h|1} \end{pmatrix} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(c - \mu_1)$$

and

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$

2. The random variable $Y_{21|1} \equiv (Y_{21} | Y_1 = c)$ is also normal, with the mean $\mu_{21|1}$.

Arbitrage Probability Estimator (5/5)

It follows easily immediately that the real arbitrage probability is

$$\begin{aligned} & q_{\text{est}} \\ =& \Pr(\omega) \\ =& \Pr(\exists h_0 > 1 \ni Y_{2h_0} \leq C_0 | Y_{21|1} \geq \mu_{21|1}) \Pr(Y_{21|1} \geq \mu_{21|1}) \\ & + \Pr(\exists h_0 > 1 \ni Y_{2h_0} \geq C_1 | Y_{21|1} \leq \mu_{21|1}) \Pr(Y_{21|1} \leq \mu_{21|1}) \\ =& 1 - \Pr(Y_{21|1} \geq \mu_{21|1}, Y_{22}, \ldots, Y_{2h} > C_0) \\ & - \Pr(Y_{21|1} \leq \mu_{21|1}, Y_{22}, \ldots, Y_{2h} < C_1), \end{aligned}$$

where

$$C_i = \mu_{21|1} + (-1)^i \log(1 - 0.01 b_{\text{cost}}\%) + (-1)^{1+i} \log(1 + 0.01 b_{\text{cost}}\%)$$

for $i = 0, 1$.

Arbitrage Probability Estimator Problem (1/3)

- Only one observed sample Σ_{11}, Σ_{12}, Σ_{21}, and Σ_{22}
  - The covariance matries are difficult to estimate.
- The weak case $Y_1 = (\beta'_{0r_0} y^*_{T-h+1}, \ldots, \beta'_{0r_0} y^*_T)'$,
  - the stationary property implies that the mean μ may be estimate as

$$\hat{\mu} = 1_{2h} \otimes \hat{\mu}^* \equiv 1_{2h} \otimes \left(\frac{1}{T} \sum_{t=1}^{T} \beta_{0r_0} y_t \right),$$

  - the (i, j)-element of the covariance matrix may be estimated as

$$\hat{\Sigma}(i,j) = \frac{1}{T - |i-j|} \sum_{t=|i-j|+1}^{T} (y_t - \hat{\mu}^*)(y_{t-|i-j|} - \hat{\mu}^*).$$

Arbitrage Probability Estimator Problem (2/3)

Definition: Insurance Principle [13]

With all risk sources independent, and with investment spread across many securities, exposure to any particular source of risk is negligible. This is an application of the law of large numbers. The reduction of risk to very low levels because of independent risk sources is called the insurance principle.

This insurance principle do not suggest that investor builds two or more portfolios based on same assets.

How can we do?

Arbitrage Probability Estimator Problem (3/3)

1. A "nice" mean-reverting spread with high variance [82],
2. Give a K-dimensional USSC process y_t with covariance matrix Σ_{y_t} is cointegrated of rank r,
3. The optimal portfolio weight $\beta_* = \beta_0 w$ for some w in $\mathbb{R}^K$ are the solutions to the following quadratic programming:

$$\beta_* = \arg \max_{\|\beta_0 w\|_1 = 1} w' \Sigma_{\beta_0' y_0} w.$$

Data

Your premium subscriptions

Product	S&P 500 One Minute Bars
License	Single User
Billing Cycle	Annual
Price	$1,000.00
Start Date	Sep 16 2016
Next Billing Cycle	Sep 16 2017
Status	Active
Options	UNSUBSCRIBE

The USSC strategy of pa minute closing prices

251.$\bar{6}$ trading days per yea

AM 09:30 to PM 04:00

are 2265 trading days with 390 trading minutes per day.

Static Setting

We suppose the number of tickers with all minute closing prices at the tested day, d, is S. The procedure and trading parameters are given as follows:

- the closed structural change interval Π is $[0.15, 0.85]$,
- the sample size T is 100,
- the number $N_{\text{criterion}}$ and top $p_{\text{criterion}}$ percentage of picked pairs in the first-stage Bernoulli sampling process are $\lceil 0.025S \rceil$ and 0.05, respectively,
- the number $N_{\text{portfolios}}$ and top p_{CI} percentage of picked pairs in the second-stage Bernoulli sampling process are 5 and 0.01, respectively,
- the predetermined successful probability s, significance level, and predetermined threshold α_{obj} are all 0.1,
- the bounded order M of the VAR model is 3,
- the threshold number of refinements is 5,
- the chosen level of b_{cost} basis points per trade per dollar of transaction cost is 10, and
- the trading time bound h is 15 (minutes).

Empirical Prediction (1/3)

For a pair $n = 1, \ldots,$ the estimated arbitrage probability $q_{\text{est},n}$ and corresponding mean-reversion outcome $q_{\text{outcome},n}$ are formulated as a regression model with a binary response variable,

$$q_{\text{outcome},n} = q_{\text{est},n} + \epsilon_n, \;\; n = 1, \ldots,$$

where ϵ_n is the error term.

Empirical Prediction (2/3)

Theorem: Khinchine's Theorem [73]

Let $\{x_t\}$ be a sequence of i.i.d. random variables with $E(x_t) = \mu < \infty$. Then

$$\bar{x}_T := \frac{1}{T}\sum_{t=1}^{T} x_t \xrightarrow{p} \mu.$$

If $\epsilon_1, \epsilon_2, \ldots$ are independent and identically distributed (i.i.d.) random variables, we have the following result:

$$\bar{\epsilon}_N \equiv \sum_{n=1}^{N} \frac{\hat{\epsilon}_n}{N} \xrightarrow{p} \mu_\epsilon \text{ as } N \to \infty,$$

where $\hat{\epsilon}_n$ is the observed error term, and μ_ϵ is the mean of ϵ_n.

Empirical Prediction (3/3)

We uniformly sample 258 start trading times and 360 trading minutes, between 16 and 375, to find that

1. the number of observed USSC pairs is 379,
2. the observed $\bar{\epsilon}_{379}$ is 0.4399, and
3. the sample mean and sample standard deviation of $q_{\text{outcome},n}$ are 0.5303 and 0.5133, respectively.

Moreover, for the estimated arbitrage probability greater than 0.0001, we have that

1. the number of observed USSC pairs is 119,
2. the observed $\bar{\epsilon}_{119}$ is 0.8358, and
3. the sample mean and sample standard deviation of $q_{\text{outcome},n}$ are 1.0000 and 0.1595, respectively.

Therefore, we underestimated the real arbitrage probability.

Conclusions

1. We proposes an automatic search procedure for pairs trading based on the cointegration approach, with an unknown single structural change,
2. As compared with other studies, the novelty of this procedure is its multi-variability and flexibility,
3. The stationary and normal properties of the generated profits, from searched pairs, imply a computable arbitrage probability,
4. The empirical result shows that the real arbitrage probability is not less than the estimated arbitrage probability.


That's all for today
今回はここまで

References (1/18)

[1] V. Abramov, M. K. Khan, and R. A. Khan. A probabilistic analysis of the trading the line strategy. *Quantitative Finance*, 8(5):499–512, aug 2008.

[2] S. K. Ahn and G. C. Reinsel. Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, 85(411):813–823, sep 1990.

[3] H. Akaike. Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, 22(2):203–217, 1970.

[4] H. Akaike. Autoregressive model fitting for control. *Annals of the Institute of Statistical Mathematics*, 23(2):163–180, 1971.

References (2/18)

[5] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *The Second International Symposium on Information Theory*, pages 267–281, Budapest, HU, 1973. Akadémiai Kiadó.

[6] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.

[7] T. W. Anderson. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, 22(3):327–351, 1951.

[8] D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, jul 1993.

References (3/18)

[9] D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point: a corrigendum. *Econometrica*, 71(1):395–397, jan 2003.

[10] D. W. K. Andrews, I. Lee, and W. Ploberger. Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, 70(1):9–38, jan 1996.

[11] J. Bai. Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4):551–563, nov 1997.

[12] L. Bauwens, G. Koop, D. Korobilis, and J. V. Rombouts. The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, 30(4):596–620, jun 2015.

References (4/18)

[13] Z. Bodie, A. Kane, and A. J. Marcus. *Essentials of Investments*. McGraw-Hill/Irwin, New York, NY, 9 edition, 2012.

[14] T. Bogomolov. Pairs trading based on statistical variability of the spread process. *Quantitative Finance*, 13(9):1411–1430, sep 2013.

[15] H. P. Boswijk. Identifiability of cointegrated systems. *Tinbergen Institute discussion paper*, (TI 95-78):1–21, 1995.

[16] H. P. Boswijk and J. A. Doornik. Identifying, estimating and testing restricted cointegrated systems: an overview. *Statistica Neerlandica*, 58(4):440–465, nov 2004.

[17] R. L. Brown, J. Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, 37(2):149–192, 1975.

References (5/18)

[18] B. Chen and Y. Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183, 2012.

[19] M. C. Chiu and H. Y. Wong. Mean-variance portfolio selection of cointegrated assets. *Journal of Economic Dynamics and Control*, 35(8):1369–1385, aug 2011.

[20] M. C. Chiu and H. Y. Wong. Optimal investment for an insurer with cointegrated assets: CRRA utility. *Insurance: Mathematics and Economics*, 52(1):52–64, jan 2013.

[21] G. C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605, 1960.

[22] N. R. Draper and H. Smith. *Applied Regression Analysis*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 3 edition, 1998.

References (6/18)

[23] J.-M. Dufour. Recursive stability analysis of linear regression relationships. *Journal of Econometrics*, 19(1):31–76, may 1982.

[24] R. J. Elliott, J. Van Der Hoek, and W. P. Malcolm. Pairs trading. *Quantitative Finance*, 5(3):271–276, jun 2005.

[25] R. F. Engle and C. W. J. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2):251–276, mar 1987.

[26] J.-P. Fouque, G. Papanicolaou, R. Sircar, and K. Sølna. *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivaties*. Cambridge University Press, New York, 2011.

[27] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst. Pairs trading: performance of a relative-value arbitrage rule. *Review of Financial Studies*, 19(3):797–827, 2006.

References (7/18)

[28] M. Gisser and T. H. Goodwin. Crude oil and the macroeconomy: tests of some popular notions: note. *Journal of Money, Credit and Banking*, 18(1):95–103, feb 1986.

[29] J. Groß. *Linear Regression*. Springer Berlin Heidelberg, German, 2003.

[30] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, 41(2):190–195, 1979.

[31] P. R. Hansen. Structural changes in the cointegrated vector autoregressive model. *Journal of Econometrics*, 114(2):261–295, jun 2003.

[32] R. V. Hogg, J. McKean, and A. T. Craig. *Introduction to Mathematical Statistics*. Pearson, New York, 7 edition, 2012.

References (8/18)

[33] R. V. Hogg and E. A. Tanis. *Probability and Statistical Inference*. Pearson, New Jersey, 8 edition, 2010.

[34] N. Huck. The high sensitivity of pairs trading returns. *Applied Economics Letters*, 20(14):1301–1304, sep 2013.

[35] N. Huck and K. Afawubo. Pairs trading and selection methods: is cointegration superior? *Applied Economics*, 47(6):599–613, feb 2015.

[36] H. Jacobs and M. Weber. On the determinants of pairs trading profitability. *Journal of Financial Markets*, 23:75–97, mar 2015.

[37] C. M. Jarque and A. K. Bera. A test for normality of observations and regression residuals. *International Statistical Review / Revue Internationale de Statistique*, 55(2):163–172, aug 1987.

References (9/18)

[38] S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254, jun 1988.

[39] S. Johansen. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, nov 1991.

[40] S. Johansen. Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics*, 54(3):383–397, aug 1992.

[41] S. Johansen. The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews*, 13(2):205–229, jan 1994.

References (10/18)

[42] S. Johansen. Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. *Journal of Econometrics*, 69(1):111–132, sep 1995.

[43] S. Johansen. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York, 1995.

[44] S. Johansen. A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica*, 70(5):1929–1961, sep 2002.

[45] S. Johansen and K. Juselius. Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52(2):169–210, may 1990.

References (11/18)

[46] S. Johansen and K. Juselius. Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics*, 53(1-3):211–244, jul 1992.

[47] G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee. *The Theory and Practice of Econometrics*. John Wiley & Sons, Inc., Canada, 2 edition, 1985.

[48] J. W. Jurek and H. Yang. Dynamic portfolio selection in arbitrage. *SSRN Electronic Journal*, 2007.

[49] P. J. Kaufman. *Trading Systems and Methods + Website*. Wiley Trading Series. John Wiley & Sons, Inc., New Jersey, 5 edition, 2013.

References (12/18)

[50] L. Kilian and U. Demiroglu. Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence. *Journal of Business & Economic Statistics*, 18(1):40–50, jan 2000.

[51] H.-J. Kim and D. Siegmund. The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, 76(3):409–423, sep 1989.

[52] C. Krauss. Statistical arbitrage pairs trading strategies: review and outlook. *Journal of Economic Surveys*, 31(2):513–545, 2017.

[53] M. Kutner, C. Nachtsheim, J. Neter, and W. Li. *Applied Linear Statistical Models*. McGraw-Hill/Irwin, New York, 5 edition, 2004.

[54] Y. Lei and J. Xu. Costly arbitrage through pairs trading. *Journal of Economic Dynamics and Control*, 56:1–19, jul 2015.

References (13/18)

[55] R. Q. Liew and Y. Wu. Pairs trading: a copula approach. *Journal of Derivatives & Hedge Funds*, 19(1):12–30, feb 2013.

[56] R. L. Lumsdaine and D. H. Papell. Multiple trend breaks and the unit-root hypothesis. *Review of Economics and Statistics*, 79(2):212–218, may 1997.

[57] H. Lütkepohl. Comparison of criteria for estimating the order of a vector autoregressive process. *Journal of Time Series Analysis*, 6(1):35–52, jan 1985.

[58] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1 edition, 2005.

[59] H. Lütkepohl, P. Saikkonen, and C. Trenkler. Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process. *The Econometrics Journal*, 4(2):287–310, dec 2001.

References (14/18)

[60] H. Lütkepohl and W. Schneider. Testing for nonnormality of autoregressive time series. *Computational Statistics Quarterly*, 5:151–168, 1989.

[61] J. G. MacKinnon and M. Ø. Nielsen. Numerical distribution functions of fractional unit root and cointegration tests. *Journal of Applied Econometrics*, 29(1):161–171, jan 2014.

[62] G. S. Maddala and I.-M. Kim. *Unit Roots Cointegration and Structural Change*. Cambridge University Press, Cambridge, 1998.

[63] N. Mantel. Why stepdown procedures in variable selection. *Technometrics*, 12(3):621–625, aug 1970.

[64] B. R. Marshall, N. H. Nguyen, and N. Visaltanachoti. ETF arbitrage: intraday evidence. *Journal of Banking and Finance*, 37(9):3486–3498, sep 2013.

References (15/18)

[65] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, New York, NY, 2006.

[66] J. Paulsen. Selected papers of hirotugu Akaike. *Journal of Time Series Analysis*, 5(2):115–127, 1984.

[67] J. Paulsen and D. Tjøstheim. On the estimation of residual variance and order in autoregressive time series. *Journal of the Royal Statistical Society*, 47(2):216–228, 1985.

[68] P. C. B. Phillips and S. N. Durlauf. Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495, 1986.

[69] P. T. Pope and J. T. Webster. The use of an F-statistic in stepwise regression procedures. *Technometrics*, 14(2):327–340, 1972.

References (16/18)

[70] R. E. Quandt. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55(290):324–330, jun 1960.

[71] B. G. Quinn. Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society*, 42(2):182–185, 1980.

[72] H. Rad, R. K. Y. Low, and R. Faff. The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance*, 16(10):1541–1558, oct 2016.

[73] C. R. Rao. *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc., New York, 2 edition, 1973.

[74] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

References (17/18)

[75] C. A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, jan 1980.

[76] J. H. Stock. Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, 55(5):1035–1056, sep 1987.

[77] C.-W. Su, L. Lu, H.-L. Chang, and Y. Zhao. Reserve accumulation and money supply in China: a time-varying analysis. *Global Economic Review*, 46(1):47–64, jan 2017.

[78] Z. Sun and B. Sun. Impact of monetary supply on Chinese nonferrous metal price movement. *Asian Economic Journal*, 31(1):17–37, mar 2017.

[79] R. S. Tsay. Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7(1):1–20, jan 1988.

References (18/18)

[80] H. J. Turtle and C. Zhang. Structural breaks and portfolio performance in global equity markets. *Quantitative Finance*, 15(6):909–922, jun 2015.

[81] A. Ulucan. An analysis of mean-variance portfolio selection with varying holding periods. *Applied Economics*, 39(11):1399–1407, jun 2007.

[82] G. Vidyamurthy. *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons, Inc., New Jersey, 2004.