Pairs Trading, Convergence Trading, Cointegration

Daniel Herlemont

email:dherlemont@yats.com - *YATS Finances & Technologies* - tel:+33 (0) 5 62 71 22 84

"Trying to model the complex interdependencies between financial assets with so restrictive concept of correlation is like trying to surf the internet with an IBM AT."
Carol Alexander

Contents

1 Introduction   **2**
    1.1   Stationary and non-stationary variables   3

2 Pairs Trading Model   **5**
    2.1   Strategy   5
    2.2   Testing for the mean reversion   6
    2.3   Screening Pairs   7
    2.4   Trading rules   7
    2.5   Risk control   8
    2.6   Risks   9
    2.7   General Discussion on pairs trading   9

3 Optimal Convergence Trading   **11**
    3.1   Optimal Trading in presence of mean reversion   11
    3.2   Continuous version of mean reversion : the Ornstein-Uhlenbeck process . . .   12

4 Granger causality   **13**

5 spurious regression   **13**

1   Introduction

Just like a a drunk man leaving a bar follows a random walk. His dog also follows a random walk on its own. The paths will diverge ... Then they go into a park where dogs are not allowed to be untied. Therefore the drunk man puts a strap on his dog and both enter into the park. Now, they share some common direction, their paths are co-integrated ... see Murray [11].

A good intro is also given by Carol Alexander in [2] "Cointegration and asset allocation: A new active hedge fund strategy".

Definition: Two time series x_t and y_t are co-integrated if, and only if, each is $I(1)$ and a linear combination $X_t - \alpha - \beta Y_t$, where $\beta_1 \neq 0$ is $I(0)$

In general, linear combinations of $I(1)$ time series are also $I(1)$. Co-integration is a particular feature not displayed between arbitrary pairs of time series.

If two time series are co-integrated, then the co-integrating vector (β_1) is unique

Granger (1981) introduced the case

$$y_t = \alpha + \beta x_t + u_t \tag{1}$$

where the individual time series are I(1) but the error term, u_t, is I(0). That is, the error term might be autocorrelated but, because it is stationary, the relationship will keep returning to the equilibrium or long-run equation $y_t = \alpha + \beta x_t$

Granger (1981) and Engle and Granger (1987) demonstrated that, if a vector of time series is cointegrated, the long-run parameters can be estimated directly without specifying the dynamics because, in statistical terms, the estimated long-run parameter estimates converge to their true values more quickly than those operating on stationary variables. That is they are super-consistent and a two-step procedure of first estimating the long-run relationship and estimating the dynamics, conditional on the long run becomes possible.

As a result, simple static models came back in vogue in the late 1980's but it rapidly became apparent that small sample biases can indeed be large (Banerjee et al, 1986).

Two major problems typically arise in a regression such as (1). First, it is not always clear whether one should regress yt on xt or vice versa. Endogeneity is not an issue asymptotically because the simultaneous equations bias is of a lower order of importance and, indeed, is dominated by the non-stationarity of the regressor. However, least squares is affected by the chosen normalisation and the estimate of one regression is not the inverse of that in the alternative ordering unless $R^2 = 1$.

Secondly, the coefficient $\hat{\beta}$ is not asymptotically normal when xt is I(1) without drift, even if ut is iid. Of course, autocorrelation in the residuals produces a bias in the least squares standard errors, even when the regressor is non-stationary, and this effect is in addition to that caused by non-stationarity.

The preceding discussion is based on the assumption that the disturbances are stationary. In practice, it is necessary to pre-test this assumption. Engle and Granger suggested a number of alternative tests but that which emerged as the popular method is the use of an ADF test on the residuals without including a constant or a time trend.

1.1 Stationary and non-stationary variables

Consider :

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{2}$$

If $|\rho| < 1$ then the series is stationary (around 0), if $|\rho| = 1$ then it is non-stationary (a random walk in this case).

A stationary series is one for which :

  (i) the mean is constant

 (ii) the variance is constant, and

(iii) $Covariance(y_t, y_{t-s})$ depends only upon the lag length s.

Strictly, this is "weak" or "second order" stationarity but is good enough for practical purposes.

We can more generally write

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t \tag{3}$$

which is stationary around $\alpha/(1 - \rho)$. (To see this, set $E(y_t) = E(y_{t-1})$ hence $E(y) = \alpha + \rho E(y) + 0$, hence $E(y) = \alpha/(1 - \rho)$.)

If $\rho = 1$, we have a random walk with drift.

We can also have a second order AR process, e.g.

$$y_t = \rho_1 y_{t-1} - \rho_2 y_{t-2} + \varepsilon_t \tag{4}$$

The conditions for stationarity of this series are given later.

We can also have a time trend incorporated:

$$y_t = \beta_1 + \beta_2 t + \rho y_{t-1} + \varepsilon_t \tag{5}$$

This will be stationary (around $\beta_1 + \beta_2 t$) if $\rho < 1$. This is called a trend stationary series (TSS).

As trend stationary series can look similar to non-stationary series.

This is unfortunate, since we should de-trend the former (take residuals from a regression on time) and difference the latter (the latter are also known as difference stationary series for this reason. It is also known as a stochastic trend). Doing the wrong transformation leads to biased estimates in regression, so it's important (but unfortunately difficult) to tell the difference.

Note that, for a non-stationary process

$$
\begin{aligned}
y_t &= \alpha + y_{t-1} + \varepsilon_t \ we \ can \ write: \\
y_0 &= 0 \ (say) \tag{6} \\
y_1 &= \alpha + 0 + \varepsilon_1 \tag{7} \\
y_2 &= \alpha + y_1 + \varepsilon_2 = 2\alpha + \varepsilon_1 + \varepsilon_2 \tag{8} \\
&\ldots \tag{9} \\
y_t &= \alpha t + \sum \varepsilon \tag{10}
\end{aligned}
$$

This implies that

$$var(y) = tvar(varepsilon) = t\sigma^2 \tag{11}$$

which tends to infinity as the sample size increases.

For a Trend Stationary Series with $\rho = 0$:

$$
\begin{aligned}
y_t &= \beta_1 + \beta_2 t + \varepsilon_t \\
y_0 &= 0 \ (say) \\
y_1 &= \beta_1 + \beta_2 + \varepsilon_1 \\
y_2 &= \beta_1 + \beta_2 2 + \varepsilon_2 \\
&\ldots
\end{aligned}
$$

(12)
(13)
(14)
(15)

$$y_t = \beta_1 + \beta_2 t + \varepsilon_t \tag{16}$$

Note the similarity of 10 and 16, apart from the nature of the error term, i.e. a Difference Stationary Series can be written as a function of t, like a Trend Stationary Series , but with a MA error term.

In finite samples a Trend Stationary Series can be approximated arbitrarily well by a Difference Stationary Series , and vice versa. The finite sample distributions are very close together and it can be hard to tell them apart.

A difference between them is that a shock to the system ($\Delta \varepsilon$) has a temporary effect upon a Trend Series but a permanent effect upon a Difference Series. If we interpret 'shock' as a change in government policy, then we can see the importance of finding whether variables are Difference Series or Trend Series.

A non-stationary series is said to be integrated, with the order of integration being the number of times the series needs to be differenced before it becomes stationary. A stationary series is integrated of order zero, I(0). For the random walk model, y I(1).

Most economic variables are $I(0)$ or $I(1)$. Interest rates are likely to be $I(0)$, they are not trended. Output, the price level, investment, etc. are likely to be $I(1)$. Some variables may be $I(2)$. Transforming to logs may affect the order of integration.

2 Pairs Trading Model

[1]

2.1 Strategy

The investment strategy we aim at implementing is a market neutral long/short strategy. This implies that we will try to find shares with similar betas, where we believe one stock

---

[1] source: from Andrei Simonov, no longer available online

will outperform the other one in the short term. By simultaneously taking both a long and short position the beta of the pair equals zero and the performance generated equals alpha.

Needless to mention, is that the hard part of this strategy is to find market neutral positions that will deviate in returns. To do this we can use a statistical tool developed by Schroder Salomon Smith Barney (SSSB). The starting point of this strategy is that stocks that have historically had the same trading patterns (i.e. constant price ratio) will have so in the future as well. If there is a deviation from the historical mean, this creates a trading opportunity, which can be exploited. Gains are earned when the price relationship is restored. The historical calculations of betas and the millions of tests executed are done by SSSB, but it is our job as portfolio managers to interpret the signals and execute the trades.

Summary:

- find two stocks prices of which have historically moved together,

- mean reversion in the ratio of the prices, correlation is not key

- Gains earned when the historical price relationship is restored

- Free resources invested in risk-free interest rate

2.2   Testing for the mean reversion

The challenge in this strategy is identifying stocks that tend to move together and therefore make potential pairs. Our aim is to identify pairs of stocks with mean-reverting relative prices. To find out if two stocks are mean-reverting the test conducted is the Dickey-Fuller test of the log ratio of the pair. In the

A Dickey-Fuller test for determining stationarity in the log-ratio $y_t = \log A_t - log B_t$ of share prices A and B

$$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t \tag{17}$$

In other words, we are regressing Δy_t on lagged values of y_t.

the null hypothesis is that $\gamma = 0$, which means that the process is not mean reverting.

If the null hypothesis can be rejected on the 99% confidence level the price ratio is following a weak stationary process and is thereby mean-reverting. Research has shown that if the confidence level is relaxed, the pairs do not mean-revert good enough to generate satisfactory returns. This implies that a very large number of regressions will be run to identify the pairs. If you have 200 stocks, you will have to run 19 900 regressions, which makes this quite computer-power and time consuming.

Schroder Salomon Smith Barney provide such calculation

2.3   Screening Pairs

By conducting this procedure, a large number of pairs will be generated. The problem is that all of them do not have the same or similar betas, which makes it difficult for us to stay market neutral. Therefore a trading rule is introduced regarding the spread of betas within a pair. The beta spread must be no larger than 0.2, in order for a trade to be executed. The betas are measured on a two-year rolling window on daily data.

We now have mean-reverting pairs with a limited beta spread, but to further eliminate the risk we also want to stay sector neutral. This implies that we only want to open a position in a pair that is within the same sector. Due to the different volatility within different sectors, we expect sectors showing high volatility to produce very few pairs, while sectors with low volatility to generate more pairs. Another factor influencing the number of pairs generated is the homogeneity of the sector. A sector like Commercial services is expected to generate very few pairs, but Financials on the other hand should give many trading opportunities. The reason why, is that companies within the Financial sector have more homogenous operations and earnings.

2.4   Trading rules

The screening process described gives us a large set of pairs that are both market and sector neutral, which can be used to take positions. This should not be done randomly, since timing is an important issue. We will therefore introduce several trading execution rules.

All the calculations described above will be updated on a daily basis. However, we will not have to do this ourselves, but we will be provided with updated numbers every day, showing pairs that are likely mean revert within the next couple of weeks. In order to execute the strategy we need a couple of trading rules to follow, i.e. to clarify when to open and when to close a trade. Our basic rule will be to open a position when the ratio of two share prices hits the 2 rolling standard deviation and close it when the ratio returns to the mean. However, we do not want to open a position in a pair with a spread that is wide and getting wider. This can partly be avoided by the following procedure: We actually want to open a position when the price ratio deviates with more than two standard deviations from the 130 days rolling mean. The position is not opened when the ratio breaks the two-standard-deviations limit for the first time, but rather when it crosses it to revert to the mean again. You can say that we have an open position when the pair is on its way back again (see the picture below).

summary:

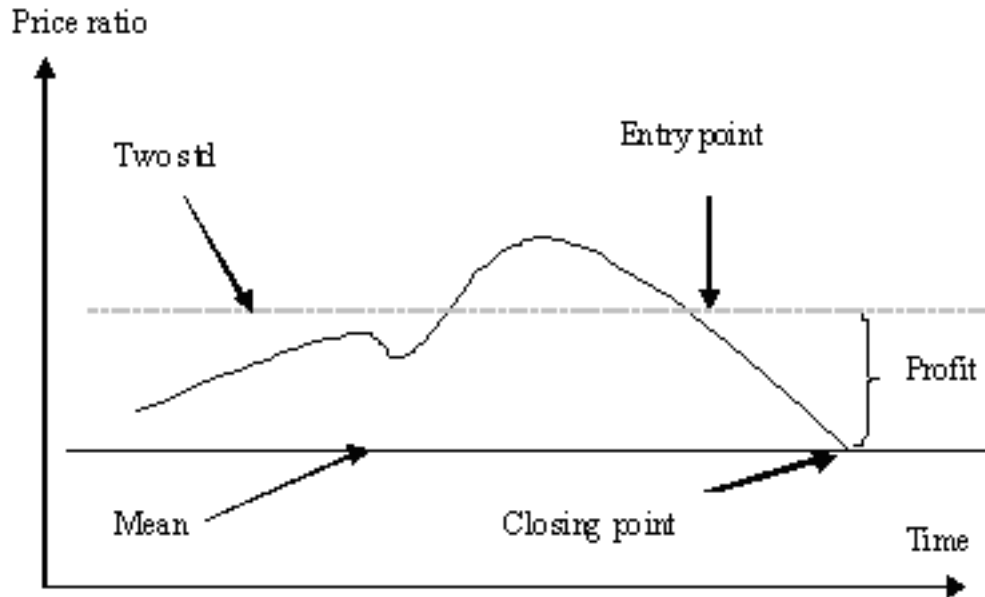- Open position when the ratio hits the 2 standard deviation band for two consecutive times

Figure 1: Pairs Trading rules

- Close position when the ratio hits the mean

2.5   Risk control

Furthermore, there will be some additional rules to prevent us from loosing too much money on one single trade. If the ratio develops in an unfavourable way, we will use a stop-loss and close the position as we have lost 20% of the initial size of the position. Finally, we will never keep a position for more that 50 days. On average, the mean reversion will occur in approximately 35 days , and there is no reason to wait for a pair to revert fully, if there is very little return to be earned. The potential return to be earned must always be higher than the return earned on the benchmark or in the fixed income market. The maximum holding period of a position is therefore set to 50 days. This should be enough time for the pairs to revert, but also a short enough time not to loose time value.

The rules described are totally based on statistics and predetermined numbers. In addition, there is a possibility for us to make our own decisions. If we for example are aware of fundamentals that are not taken into account in the calculations and that indicates that

there will be no mean reversion for a specific pairs, we can of course avoid investing in such pairs.

From the rules it can be concluded that we will open our last position no later than 50 days before the trading game ends. The last 50 days we will spend trying to close the trades at the most optimal points of time.

Summary:

- Stop loss at 20% of position value

- Beta spread < 0.2

- Sector neutrality

- Maximum holding period < 50 trading days

- 10 equally weighted positions

2.6   Risks

As already mentioned, through this strategy we do almost totally avoid the systematic market risk. The reason there is still some market risk exposure, is that a minor beta spread is allowed for. In order to find a sufficient number of pairs, we have to accept this beta spread, but the spread is so small that in practise the market risk we are exposed to is ignorable. Also the industry risk is eliminated, since we are only investing in pairs belonging to the same industry.

The main risk we are being exposed to is then the risk of stock specific events, that is the risk of fundamental changes implying that the prices may never mean revert again, or at least not within 50 days. In order to control for this risk we use the rules of stop-loss and maximum holding period. This risk is further reduced through diversification, which is obtained by simultaneously investing in several pairs. Initially we plan to open approximately 10 different positions. Finally, we do face the risk that the trading game does not last long enough. It might be the case that our strategy is successful in the long run, but that a few short run failures will ruin our overall excess return possibilities.

2.7   General Discussion on pairs trading

There are generally two types of pairs trading: statistical arbitrage convergence/divergence trades, and fundamentally-driven valuation trades. In the former, the driving force for the trade is a aberration in the long-term spread between the two securities, and to realize the

mean-reversion back to the norm, you short one and go long the other. The trick is creating a program to find the pairs, and for the relationship to hold.

The other form of pairs trading would be more fundamentally-driven variation, which is the purvey of most market-neutral hedge funds: in essence they short the most overvalued stock(s) and go long the undervalued stock(s). It's normal to "pair" up stocks by having the same number per sector on the long and short side, although the traditional "pairs" aren't used anymore. Pairs trading had originally been the domain of BD's in the late 70's, early 80's before it dissipated somewhat due to the bull market (who would want to be market-neutral in a rampant bull market), and the impossibility of assigning "pairs" due to the morphing of traditional sectors and constituents. Most people don't perform traditional "pairs trading" anymore (i.e. the selection of two similar, but mispriced, stocks from the same industry/sector), but perform a variation. Goetzmann et al wrote a paper on it a few years back, but at the last firm I worked at, the research analyst "pooh-paahed" it because he couldn't get the same results: he thinks Goetzmann [7] either waived commissions, or worse, totally ignored slippage (i.e. always took the best price, not the realistically one). Here's the paper : source: forum http://www.wilmott.com

some quotations from this paper: "take a longshort position when they diverge." A test requires that both of these steps must be parameterized in some way. How do you identify "stocks that move together?" Need they be in the same industry? Should they only be liquid stocks? How far do they have to diverge before a position is put on? When is a position unwound? We have made some straightforward choices about each of these questions. We put positions on at a twostandard deviation spread, which might not always cover transactions costs even when stock prices converge. Although it is tempting to try potentially more profitable schemes, the danger in datasnooping refinements outweigh the potential insights gained about the higher profits that could result from learning through testing. As it stands now, datasnooping is a serious concern in our study. Pairs trading is closely related to a widely studied subject in the academic literature mean reversion in stock prices. 2 We consider the possibility that we have simply reformulated a test of the previously documented tendency of stocks to revert towards their mean at certain horizons. To address this issue, we develop a bootstrapping test based upon random pair choice. If pairstrading profits were simply due to meanreversion, then we should find that randomly chosen pairs generate profits, i.e. that buying losers and selling winners in general makes money. This simple contrarian strategy is unprofitable over the period that we study, suggesting that mean reversion is not the whole story. Although the effect we document is not merely an extension of previously known anomalies, it is still not immune to the datasnooping argument. Indeed we have explicitly "snooped" the data to the extent that we are testing a strategy we know to have been actively exploited by risk arbitrageurs. As a consequence we cannot be sure that past trading profits under our simple strategies will continue in the future. This potential

critique has another side, however. The fact that pairs trading is already wellknown risk-arbitrage strategy means that we can simply test current practice rather than develop our filterrule ad hoc.

3 Optimal Convergence Trading

From [10], Vladislav Kargin: "Consider an investment in a mispriced asset. An investor can expect that the mispricing will be eliminated in the future and play on the convergence of the asset price to its true value. This play is risky because the convergence is not immediate and its exact date is uncertain. Often both the expected benefit and leveraging his positions, that is, by borrowing additional investment funds. An important question for the investor is what is the optimal leverage policy. There are two intuitively appealing strategies in this situation. The first one is to invest only if the mispricing exceeds a threshold and to keep the position unchanged until the mispricing falls below another threshold (an (s, S) strategy). For this strategy the relevant questions are what are the optimal thresholds and what are the properties of the investment portfolio corresponding to this strategy. The second type of strategy is to continuously change positions according to the level of mispricing. In this case, we are interested in the optimal functional form of the dependence of the position on the mispricing."

See also discussion on optimal growth strategies in [9].

3.1 Optimal Trading in presence of mean reversion

In [14], Thompson define close form of trading threshold strategies in presence of Ornstein-Uhlenbeck process and fixed transaction costs c.

if the price of the OU process is

$$dS_t = \sigma dB_t - \gamma S_t dt$$

The optimal strategy is a threshold strategy, ie to buy if $S_t \leq -b/\gamma$ and to sell $S_t \geq -b/\gamma$ whre b satisfies:

$$2b - \gamma c = 2e^{-b^2/(\gamma\sigma^2} \int_0^b e^{u^2/(\gamma\sigma^2} du$$

with c fixed transaction cost.

3.2   Continuous version of mean reversion : the Ornstein-Uhlenbeck process

Suppose that the dynamic of the mis-pricing can be modelled by an AR(1) process. RA(1) process is the discrete-time counterpart to the Ornstein-Uhlenbeck (OU) process in continuous time.

$$dX_t = \beta(\alpha - X_t)dt + \sigma dW_t \tag{18}$$

where W_t is a standard Wiener process, $\sigma > 0$ and α, β are constants. So the X_t process drifts towards α.

OU process also has a normal transition density function given by :

$$f(X_t = x, t; X_{t_0} = x_0, t_0) = \frac{1}{\sqrt{2\pi s^2(t)}} e^{-\frac{(x-m(t))^2}{2s^2(t)}} \tag{19}$$

with mean

$$m(t) = \alpha + (x_0 - \alpha)e^{-\beta(t-t_0)} \tag{20}$$

with variance:

$$s^(t) = \frac{\sigma^2}{2\beta}[1 - e^{-2\beta(t-t_0)}] \tag{21}$$

If the process displays the property of mean reversion ($\beta > 0$), then as $t_0 \to -\infty$ or $t - t_0 \to +\infty$, the marginal density of the process is invariant to time, ie OU process is stationary in the strict sense.

See that there is a time decay term for the variance. For long time horizon the variance of this process tends to $\sigma^2/2\beta$. So, unlike the Brownian motion the variance is bounded (not grows to infinite).

The equation describing dX_t, the arithmetic Ornstein-Uhlenbeck equation presented above is a continuous time version of the first-order autoregressive process, AR(1), in discrete time. It is the limiting case (Δt tends to zero) of the AR(1) process:

$$x_t - x_{t-1} = \alpha(1 - e^{-\beta}) + (e^{-\beta} - 1)x_t - 1 + \varepsilon_t \tag{22}$$

Where ε_t is normally distributed with mean zero and standard deviation

$$\sigma_\varepsilon = 1 - e^{-2\beta}\frac{\sigma^2}{2\beta} \tag{23}$$

In order to estimate the parameters of mean-reversion, run the regression:

$$x_t - x_{t-1} = a + bx_{t-1} + \varepsilon_t \tag{24}$$

Calculate the parameters:

$$\begin{aligned} \alpha &= -a/b; \\ \beta &= -ln(1+b); \end{aligned} \tag{25}$$

$$\sigma = \sigma_\varepsilon \sqrt{\frac{ln(1+b)}{(1+b)^2 - 1}} \tag{26}$$

The choice of the representation may depend on the data. For example, with daily data, the AR discrete form is preferable, with high frequency data, it might preferable to use the continuous time model.

One important distinction between random walk and stationary AR(1) processes: for the last one all the shocks are transitory, whereas for random walk all shocks are permanent

Mean-Reversion Combined with Exponential Drift

It is possible combine Geometric Brownian Motion (exponential drift) with mean-reverting model.

$$\frac{dX}{X} = \left(\alpha + \eta(\hat{X}e^{\alpha t} - X)\right) dt + \sigma dW \tag{27}$$

4 Granger causality

According to Granger (1981), a times series, Xt, is said to cause another times series, Yt, if present Y can be predicted better by using the value of X. The first step in the empirical analysis is to examine the stationarity of the price series.

5 spurious regression

We now look at situations where the validity of the linear regression model is dubious - where variables are trended or, more formally, non-stationary (not quite the same thing). Regressions can be spurious when variables are n-s, i.e. you appear to have 'significant' results when in fact you haven't.

Nelson and Plosser ran an experiment. They generated two random variables (these are random walks):

$$\begin{aligned} x_t &= x_{t-1} + \varepsilon_t \\ y_t &= y_{t-1} + \nu_t \end{aligned} \tag{28}$$

where both errors have the classical properties and are independent. y and x should therefore be independent.

Regressing y on x, N & P got a 'significant' result (at the 5% level) 75% of the time !!! This is worrying! This is a spurious regression and it occurs because of the common trends in y and x.

In these circumstances, the t and F statistics do not have the standard distributions. Unfortunately, the problem generally gets worse with a larger sample size.

Such problems tend to occur with non-stationary variables and, unfortunately, many economic variables are like this.

6   Unit Root Hypohtesis Testing

6.1   Dickey Fuller Tests

Formally, a stochastic process is stationary if all the roots of its characteristic equation are ¿1 in absolute value. Solving is the same as solving a difference equation.

Examples For

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{29}$$

we rewrite it as $y_t - \rho y_{t-1} = \varepsilon_t$ or $(1 - \rho L)y_t = \varepsilon_t$. Hence this will be stationary if the root of the characteristic equation $1 - \rho L = 0$ is > 1. The root is $L = 1/\rho$ which is > 1 if $\rho < 1$. This is the condition for stationarity.

Example II: $y_t = 2.8y_{t-1} - 1.6y_{t-2} + \varepsilon_t$ The characteristic equation is $1 - 2.8L + 1.6L^2 = 0$. From this we get $L = 1.25$ and $L = 0.5$ are the roots. Both roots need to be greater than 1 in absolute value for stationarity. This does not apply here, so the process is non-stationary.

However, in practice we do not know the ρ values, we have to estimate them and then test whether the roots are all > 1. We could estimate

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{30}$$

and test

$$H0: \ \rho = 1 \ (non \ stationary) \tag{31}$$

versus

$$H1: \ \rho < 1 \ (stationary) \tag{32}$$

using a t-test. Unfortunately, if $\rho = 1$, the estimate of r is biased downwards (even in large samples) and also the t-distribution is inappropriate. Hence can't use standard methods.

Instead we use the Dickey-Fuller test. Re-write 30

$$\Delta y_t = \rho^* y_{t-1} + \varepsilon_t \tag{33}$$

where $\rho^* = \rho - 1$.

Now we test

$$H0: \ \rho^* = 0 \ (non \ stationary) \tag{34}$$

versus

$$H1: \ \rho^* < 0 \ (stationary) \tag{35}$$

We cannot use critical values from the t-distribution, but D-F provide alternative tables to use.

The D-F equation only tests for first order autocorrelation of y. If the order is higher, the test is invalid and the D-F equation suffers from residual correlation. To counter this, add lagged values of Δy to the equation before testing. This gives the Augmented Dickey-Fuller test. Sufficient lags should be added to ensure ε is white noise.

95% critical value for the augmented Dickey-Fuller statistic $ADF = -3.0199$

It is important to know the order of integration of non-stationary variables, so they may be differenced before being included in a regression equation. The ADF test does this, but it should be noted that it tends to have low power (i.e. it fails to reject H0 of non-stationarity even when false) against the alternative of a stationary series with ρ near to 1.

6.2   Variants on the Dickey-Fuller Test

The Dickey-Fuller test requires that the us be uncorrelated. But suppose we have a model like the following, where the first difference of Y is a stationary AR(p) process:

$$\Delta Y_t = \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t \tag{36}$$

This model yields a model for Y_t that is:

$$Y_t = Y_{t-1} + \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t \tag{37}$$

If this is really what's going on in our series, and we estimate a standard D.F. test:

$$Y_t = \hat{\rho} Y_{t-1} + u_t \tag{38}$$

the term $\sum_{i=1}^{p} d_i \Delta Y_{t-i}$ gets lumped into the errors u_t. This induces an AR(p) structure in the us, and the standard D.F. test statistics will be wrong. There are two ways of dealing with this problem:

- Change the model (known as the augmented Dickey-Fuller test), or

- Change the test statistic (the Phillips-Perron test).

6.3   The Augmented Dickey-Fuller Test

Rather than estimating the simple model, we can instead estimate:

$$\Delta Y_t = \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t \tag{39}$$

and test whether or not $\rho = 0$. This is the Augmented Dickey-Fuller test. As with the D-F test, we can include a constant/trend term to differentiate between a series with a unit root and one with a deterministic trend.

$$Y_t = \alpha + \beta t + Y_{t-1} + \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t \tag{40}$$

The purpose of the lags of ΔY_t is to ensure that the us are white noise. This means that in choosing p (the number of lagged ΔY_{t-i} terms to include), we have to consider two things:

1. Too few lags will leave autocorrelation in the errors, while

2. Too many lags will reduce the power of the test statistic.

This suggests, as a practical matter, a couple different ways to go about determining the value of p:

1. Start with a large value of p, and reduce it if the values of di are insignificant at long lags - This is generally a pretty good approach.

2. Start with a small value of p, and increase it if values of di are significant. This is a less-good approach...

3. Estimate models with a range of values for p, and use an AIC/BIC/ F-test to determine which is the best option. This is probably the best option of all...

A sidenote: AIC and BIC tests:
The Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) are general tests for model specification. They can be applied across a range of different areas, and are like F-tests in that they allow for the testing of the relative power of nested models. Each, however, does so by penalizing models which are overspecified (i.e., those with "too many" parameters). The AIC statistic is:

$$AIC(p) = \log \sigma_p^2 + \frac{2p}{N} \tag{41}$$

where N is the number of observations in the regression, p is the number of parameters in the model (including ρ and α), and σ_p^2 is the estimated σ^2 for the regression including p total parameters.

Similarly, the BIC statistic is calculated as:

$$BIC(p) = ln\sigma_p^2 + \frac{2p\log N}{N} \tag{42}$$

2.1 The KPSS test One potential problem with all the unit root tests so far described is that they take a unit root as the null hypothesis. Kwiatkowski et. al. (1992) provide an alternative test (which has come to be known as the KPSS test) for testing the null of stationarity against the alternative of a unit root. This method considers models with constant terms, and either with or without a deterministic trend term. Thus, the KPSS test tests the null of a level- or trend-stationary process against the alternative of a unit root. Formally, the KPSS test is equal to:

estimated error variance from the regression:

$$Y_t = \alpha + \varepsilon_t \tag{43}$$

or:

$$Y_t = \alpha + \beta t + \varepsilon_t \tag{44}$$

for the model with a trend. The practical advantages to the KPSS test are twofold. First, they provide an alternative to the DF/ADF/PP tests in which the null hypothesis is stationarity. They are thus good "complements" for the tests we have focused on so far. A common strategy is to present results of both ADF/PP and KPSS tests, and show that the results are consistent (e.g., that the former reject the null while the latter fails to do so, or vice-versa). In cases where the two tests diverge (e.g., both fail to reject the null), the possibility of "fractional integration" should be considered (e.g. Baillie 1989; Box-Steffensmeier and Smith 1996, 1998).

General Issues in Unit Root Testing The Sims (1988) article I assigned is to point out an issue with unit root econometrics in general: that classicists and Bayesians have very different ideas about the value of knife-edge unit root tests like the ones here.1 Unlike classical statisticians, Bayesians regard (the "true" value of the autocorrelation parameter) as a random variable, and the goal to describe the distribution of this variable, making use of the information contained in the data. One result of this is that, unlike the classical approach (where the distribution is skewed), the Bayesian perspective allows testing using standard t distributions. For more on why this is, see the discussion in Hamilton.

Another issue has to do with lag lengths. As in the case of ARIMA models, choosing different lag lengths (e.g. in the ADF, PP and KPSS tests) can lead to different conclusions.

This is an element of subjectivity that one needs to be aware of, and sensitivity testing across numerous different lags is almost always a good idea. Finally, the whole reason we do unit root tests will become clearer when we talk about cointegration in a few weeks.

6.4 Error Correction Model

Error Correction Model (ECM) is a step forward to determine how variables are linked together.

1. Test the variables for order of integration. They must both (all) be I(d).
2. Estimate the parameters of the long run relationship. For example,

$$y_t = \alpha + \beta x_t + e_t \tag{45}$$

when y_t and z_t are cointegrated OLS is super consistent. That is, the rate of convergence is T^2 rather than just T in Chebyshev's inequality.

3. Denote the residuals from step 2 as and fit the model

$$\Delta \hat{e}_t = a\hat{e}_t + \eta_t \tag{46}$$

The null and alternate hypotheses are

$$H_0 : a = 0 \implies unit\ root = no\ cointegration$$
$$H_1 : a \neq 0 \implies no\ unit\ root = cointegration \tag{47}$$

Interpretation: Rejection of the Null implies the residual is stationary. If the residual series is stationary then y_t and x_t must be cointegrated.

4. If you reject the null in step 3 then estimate the parameters of the Error Correction Model

$$
\begin{aligned}
\Delta y_t &= \alpha_1 + \alpha_y(y_{t-1} - \beta x_{t-1} + \sum_{i=1}^{p} a_{11}^{(i)} \Delta y_{t-i} + \sum_{i=1}^{q} a_{12}^{(i)} \Delta x_{t-i} + e_{yt} \\
\Delta x_t &= \alpha_2 + \alpha_x(y_{t-1} - \beta x_{t-1} + \sum_{i=1}^{p} a_{21}^{(i)} \Delta y_{t-i} + \sum_{i=1}^{q} a_{22}^{(i)} \Delta x_{t-i} + e_{xt}
\end{aligned} \tag{48}
$$

ECM is generalized to vectors:

The components of the vector $x_t = (x1t, x2t, .., xnt)$ are cointegrated of order (d,b), denoted by xt   CI(d,b), if

All components of x_t are I(d) and There exists a vector $\beta = (\beta_1, \beta_2, ..., \beta_n)$ such that βx_t is I(d-b), where b ¿ 0. Note β is called the cointegrating vector.

Points to remember:

To make β unique we must normalize on one of the coefficients.

All variables must be cointegrated of the same order. But, all variables of the same I(d) are not necessarily cointegrated.

If x_t is $nx1$ then there may be as many as $n - 1$ cointegrating vectors. The number of cointegrating vectors is called the cointegrating rank.

An interpretation of cointegrated variables is that they share a common stochastic trend.

Given our notions of equilibrium in economics, we must conclude that the time paths of cointegrated variables are determined in part by how far we are from equilibrium. That is, if the variables wander from each other, there must be some way for them to get back together. This is the notion of error correction.

Granger Representation theorem : "Cointegration implies Error Correction Model (ECM)."

6.5   Discussions

6.5.1   Aaron at willmot.com

see original discussion link

Cointegration is covered in any good econometrics textbook. If you need more depth, Johansen's Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press, 1995, is good. I do not recommend the original Engle and Granger paper (1987).

Two series are said to be (linearly) "cointegrated" if a (linear) combination of them is stationary. The practical effect in Finance is we deal with asset prices directly instead of asset returns.

For example, suppose I want to hold a market neutral investment in stock A (I think stock A will outperform the general market, but I have no view about the direction of the overall market). Traditionally, I buy $1,000,000 of stock A and short $1,000,000 times Beta of the Index. Beta is derived from the covariance of returns between stock A and the Index. Looking at things another way, I choose the portfolio of A and the Index that would have had minimum variance of return in the past (over the estimation interval I used for Beta, and subject to the constraint that the holding of stock A is $1,000,000).

A linear cointegration approach is to select the portfolio in the past that would have been most stationary. There are a variety of ways of defining this (just as there are a variety of ways of estimating Beta) but the simplest one is to select the portfolio with the minimum extreme profit or loss over the interval. Note that the criterion is based on P&L of the portfolio (price) not return (derivative of price).

The key difference is correlation is extremely sensitive to small time deviations, cointegration is not. Leads or lags in either price reaction or data measurement make correlations useless. For example, suppose you shifted the data in the problem above, so your stock quotes were one day earlier than you Index quotes. That would completely change the Beta, probably sending it near zero, but would have little effect on the cointegration analysis.

Economists need cointegration because they deal with bad data, and their theories incorporate lots of unpredictable leads and lags. Finance, in theory, deals with perfect data with no leads or lags. If you have really good data of execution prices, cointregation throws out your most valuable (in a money-making sense) information. If you can really execute, you don't care if there's only a few seconds in which to do so. But if you have bad data, either in the sense that the time is not well-determined or that you may not be able to execute, cointegration is much safer.

In a sense, people have been using cointegration for asset management as long as they have been computing historical pro forma strategy returns and looking at the entire chart, not just the mean and standard deviation (or other assumed-stationary parameters).

My feeling is cointegration is essential for risk management and hedging, but useless for trading and pricing. Correlation is easy and well-understood. You can use it for risk-management and hedging, but only if you backtest (which essentially checks the results against a co-integration approach) to find the appropriate adjustments and estimation techniques. Correlation is useful for trading and pricing (sorry Paul) but only if you allow stochastic covariance matrices.

More formally, if a vector of time series is I(d) but a linear combination is integrated to a lower order, the time series are said to be co-integrated.

6.5.2   Allen and Fildes

from : Econometric Forecasting http://www.lums2.lancs.ac.uk/MANSCI/Staff/EconometricForecasting.pd

These are the arguments in favor of testing whether a series has a unit root:

(1) It gives information about the nature of the series that should be helpful in model specification, particularly whether to express the variable in levels or in differences.

(2) For two or more variables to be cointegrated each must possess a unit root (or more than one).

These are the arguments against testing:

(1) Unit root tests are fairly blunt tools. They have low power and often conclude that a unit root is present when in fact it is not. Therefore, the finding that a variable does not possess a unit root is a strong result. What is perhaps less well known is that many unit-root tests suffer from size distortions. The actual chance of rejecting the null hypothesis of a unit root, when it is true, is much higher than implied by the nominal significance level. These

findings are based on 15 or more Monte Carlo studies, of which Schwert (1989) is the most influential (Stock 1994, p. 2777).

(2) The testing strategy needed is quite complex.

In practice, a nonseasonal economic variable rarely has more than a single unit root and is made stationary by taking first differences. Dickey and Fuller (1979) recognized that they could test for the presence of a unit root by regressing the first-differenced series on lagged values of the original series. If a unit root is present, the coefficient on the lagged values should not differ significantly from zero. They also developed the special tables of critical values needed for the test.

Since the publication of the original unit root test there has been an avalanche of modifications, alternatives, and comparisons. Banerjee, Dolado, Galbraith, and Hendry (1993, chapter 4) give details of the more popular methods. The standard test today is the augmented Dickey-Fuller test (ADF), in which lagged dependent variables are added to the regression. This is intended to improve the properties of the disturbances, which the test requires to be independent with constant variance, but adding too many lagged variables weakens an already low-powered test.

Two problems must be solved to perform an ADF unit-root test: How many lagged variables should be used? And should the series be modeled with a constant and deterministic trend which, if present, distort the test statistics? Taking the second problem first, the ADF-GLS test proposed by Elliott, Rothenberg, and Stock (1996) has a straightforward strategy that is easy to implement and uses the same tables of critical values as the regular ADF test.

First, estimate the coefficients of an ordinary trend regression but use generalized least squares rather than ordinary least squares. Form the detrended series, y^d, given by $y_t^d = y_t - \beta_0 - \beta_1 t$, where β_0 and β_1 are the coefficients just estimated. In the second stage, conduct a unit root test with the standard ADF approach with no constant and no deterministic trend but use y^d instead of the original series.

To solve the problem of how many lagged variables to use, start with a fairly high lag order, for example, eight lags for annual, 16 for quarterly, and 24 for monthly data. Test successively shorter lags to find the length that gives the best compromise between keeping the power of the test up and keeping the desirable properties of the disturbances. Monte Carlo experiments reported by Stock (1994) and Elliott, Rothenberg and Stock (1996) favor the Schwartz BIC over a likelihood-ratio criterion but both increased the power of the unit-root test compared with using an arbitrarily fixed lag length. We suspect that this difference has little consequence in practice. Cheung and Chinn (1997) give an example of using the ADF-GLS test on US GNP.

Although the ADF-GLS test has so far been little used it does seem to have several advantages over competing unit-root tests:

(1) It has a simple strategy that avoids the need for sequential testing starting with the

most general form of ADF equation (as described by Dolado, Jenkinson, and Sosvilla-Rivero (1990, p. 225)).

(2) It performs as well as or better than other unit-root tests. Monte Carlo studies show that its size distortion (the difference between actual and nominal significance levels) is almost as good as the ADF t-test (Elliott, Rothenberg & Stock 1996; Stock 1994) and much less than the Phillips-Perron Z test (Schwert 1989). Also, the power of the ADF-GLS statistic is often much greater than that of the ADF t-test, particularly in borderline situations.

Elliott et. al (1996) showed that there is no uniformly most powerful test for this problem and derived tests that were approximately most powerful in the sense that they have asymptotic power close to the envelope of most powerful tests for this problem.

7   Variance Ratio Test

- Based on the idea that, if a series is stationary, the variance of the series is not increasing over time; while a series with a unit root has increasing variance.

- Intuition: Compare the variance of a subset of the data "early" in the series with a similarly-sized subset "later" in the process. In the limit, for a stationary series, these two values should be the same, while they will be different for an I(1) series. Thus, the null hypothesis is stationarity, as for the KPSS test.

- There's a good, brief discussion of these tests in Hamilton (p. 531-32). Other cites are Cochrane (1988), Lo and McKinlay (1988), and Cecchetti and Lam (1991).

The variance ratio methodology tests the hypothesis that the variance of multi-period returns increases linearly with time.

Hence if we calculate the variance σ^2 of a series of returns every Δt periods, the null hypothesis suggests that sampling every $k * \Delta t$ periods will lead to a variance $k\sigma^2$:

$$Variance(r_{k\Delta t}) = kVariance(r_{\Delta t}) \tag{49}$$

The variance ratio is significantly below one under mean reversion, and above one under random walk:

$$VR(k) = \frac{Variance(r_{k\Delta t})/k}{Variance(r_{\Delta t})} \quad \begin{aligned} &= 1 \; under \; random \; walk \\ &< 1 \; under \; mean \; reversion \\ &> 1 \; under \; mean \; aversion \end{aligned} \tag{50}$$
$$\tag{51}$$

More precisely, in the usual fixed k asymptotic treatment, under the null hypothesis that the x_t follow a random walk with possible drift, given by

$$x_t = \mu + x_{t-1} + \varepsilon_t \tag{52}$$

where μ is areal number and ε_t is a sequence of ε_t is a sequence of zero mean independent random variables, it is possible to show that

$$\sqrt{n}(VR(k) - 1) \to N(0, \sigma_k^2) \tag{53}$$

where σ_k^2 is some simple function of k (this is not the variance, due to overlapping observations, to make sample size sufficiently large for large k, and to correct the bias in variance estimators)

note that this result is quite general, and stands under the simple hypothesis that ε_t is a sequence of zero mean independent random variables. Any significant deviation from 7.1 means that ε_t are not independent random variables. This result extends to the case where the ε_t are a martingale difference series with conditional heteroscedasticity though the variance σ_k^2 has to be adjusted a little.

The use of the VR statistic can be advantageous when testing against several interesting alternatives to the random walk model, most notably those hypotheses associated with mean reversion. In fact, a number of authors (e.g., Lo and Mackinlay (1989), Faust (1992) and Richardson and Smith (1991)) have found that the VR statistic has optimal power against such alternatives.

Note that $VR(k)$ can be writen as :

$$
\begin{aligned}
VR(k) &= \frac{Variance(r_{kt})}{kVariance(r_t)} \\
&= \frac{Variance(r_t + r_{t+1} + ... + r_{t+k})}{kVariance(r_t)}
\end{aligned}
\tag{54}
$$

This expression can be expanded in :

$$VR(k) = 1 + 2\sum_{i=1}^{k-1}(1 - \frac{i}{k})\rho_i \tag{55}$$

where ρ_i is the i'th term in the autocorrelation function (ACF) of returns. This expression holds asymptotically. Note that this expression can be used to calculate the ACF at various lags. For example, for $k = 2$

$$VR(2) = 1 + \rho_1 \tag{56}$$

Note that if $VR(2)$ is significantly under one,is the same as as a negative autocorrelation at lag one : $\rho_1 < 0$

7.1   Implementation

Let x_i $i = 0, N$ be the series of increments (log of returns for example)

$$
\begin{aligned}
\hat{\sigma}_a^2 &= \frac{1}{N} \sum_{i=1} N (x_i - x_{i-1} - \hat{\mu})^2 \\
\hat{\sigma}_c^2 &= \frac{1}{M} \sum_{i=k} M (x_i - x_{i-k} - k\hat{\mu})^2
\end{aligned}
\tag{57}
$$

where x_k is the log price at time k, N, the sample size, $M = k(N - k + 1)(1 - k/N)$ and $\hat{\mu} = (x_N - x_0)/N$ estimate of the mean. VR(k) is defined as $\hat{\sigma}_a^2 / \hat{\sigma}_c^2$

Testing for the null hypothesis is to test if is normally distributed. Nomality classical tests can be applied, like z scores, Kolmogorov Smirnov test (or rather a Lillifors test).

8   Absolute Return Ratio Test

source: Groenendijky & Al., A Hybrid Joint Moment Ratio Test for Financial Time Series see [**?**] If one augments the martingale assumption for financial asset prices with the condition that the martingale differences have constant (conditional) variance, it follows that the variance of asset returns is directly proportional to the holding period. This property has been used to construct formal testing procedures for the martingale hypothesis, known as variance ratio tests. Variance ratio tests are especially good at detecting linear dependence in the returns. While the variance ratio statistic describes one aspect of asset returns, the idea behind this statistic can be generalized to provide a more complete characterization of asset return data. We focus on using a combination of the variance ratio statistic and the first absolute moment ratio statistic.

The first absolute moment ratio statistic by itself is useful as a measure of linear dependence if no higher order moments than the variance exist.

In combination with the variance ratio statistic it can be used to disentangle linear dependence from other deviations of the standard assumption in finance of unconditionally normally distributed returns. In particular, the absolute moment ratio statistic provides information concerning the tail of the distribution and conditional heteroskedasticity.

By using lower order moments of asset returns in the construction of volatility ratios, e.g., absolute returns, one relaxes the conditions on the number of moments that need to exist for standard asymptotic distribution theory to apply.

We formally prove that our general testing methodology can in principle even be applied for return distributions that lie in the domain of attraction of a stable law (which includes the

normal distribution as a special case). Stable laws, apart from the normal distribution, have infinite variance, such that our approach is applicable outside the finite-variance paradigm.

Since in empirical work there often exists considerable controversy about the precise nature of the asset return.

The first absolute moment has been used before as a measure of volatility,

Muller et al. observe a regularity in the absolute moment estimates which is not in line with the presumption of i.i.d. normal innovations; this regularity was labelled the scaling law. In this paper we consider the ratios of these absolute moment estimates, we obtain their statistical properties under various distributional assumptions, and we explain the observed regularity behind the 'scaling law'. In particular, we show why the deviations observed by Muller et al. should not be carelessly interpreted as evidence against the efficient market hypothesis. Furthermore, we show that the absolute moment ratio statistics contain much more information than the scaling law. Especially, when the statistic is used in combination with the variance ratio statistic, most of the characteristic features of asset returns come to the fore. Specifically, we advocate the simultaneous use of volatility statistics based on first (absolute returns) and second order moments (variances).

In such a way we construct a test which is not only suited to detect linear dependence in asset returns, but also fat-tailedness and non-linear dependence, e.g., volatility clustering.

We analytically show why moment ratios based on absolute returns can be used to detect fat-tailedness and volatility clustering, while standard variance ratios convey no information in this respect.

Discriminating between the alternative phenomena is important, since they have different implications for portfolio selection and risk management.

Throughout the paper, we rely on a convenient graphical representation of the statistics: the moment ratio curves. The formal testing procedure we propose in this paper heavily builds on the bootstrap. By performing a non-parametric bootstrap based on the empirical returns, we construct uniform confidence intervals for the range of moment ratios considered.

Absolute returns exhibits the highest correlation (Rama Cont [4]).

9   Multi variate co-integration - Vector Error Correction Modelling

Among the general class of the multivariate ARIMA (AutoRegressive Integrated Moving Average) model, the Vector Autoregressive (VAR) model turns out to be particularly convenient for empirical work. Although there are important reasons to allow also for moving average errors (e.g. L. utkepohl 1991, 1999), the VAR model has become the dominant work

horse in the analysis of multivariate time series. Furthermore, Engle and Granger (1987) show that VAR model is an attractive starting point to study the long run relationship between time series that are stationary in first differences. Since Johansen's (1988) seminal paper, the co-integrated VAR model has become very popular in empirical macroeconomics.

see resources.

10   Resources

- [6], Engle and Granger seminal paper: "Cointegration and Error-Correction: Representation, Estimation and Testing",

- ***** Explaining Cointegration Analysis: David F. Hendry and Katarina Juselius: part I, cached part II, cached

- Carol Alexander is specialized in cointegration trading and index tracking, "Cointegration and asset allocation: A new active hedge fund strategy" [2] includes a good intro to cointegration, see also, http://www.bankingmm.com and related paper of Alexander [1] "Cointegration-based trading strategies: A new approach to enhanced index tracking and statistical arbitrage"

- In "Intraday Price Formation in US Equity Index Markets" [8], Joel Hasbrouck is studying relationships between stocks and futures and ETF. including implementation source codes, cached and presentation slides *****

- Chambers [3] This paper analyses the effects of sampling frequency on the properties of spectral regression estimators of cointegrating parameters.

- "Numerically Stable Cointegration Analysis" [5] is a practical impementation to estimate commen trends: Cointegration analysis involves the solution of a generalized eigenproblem involving moment matrices and inverted moment matrices. These formulae are unsuitable for actual computations because the condition numbers of the resulting matrices are unnecessarily increased. Our note discusses how to use the structure of the problem to achieve numerically stable computations, based on QR and singular value decompositions.

- [11], a simple illustration of cointegration with a drunk man and his dog ... ****

- [13] Adrian Trapletti paper on intraday cointegration for forex.

- Common stochastic trends, cycles and sectoral fluctuations cached *****

- johansen test critical values

- web cached This paper uses Reuters high-frequency exchange rate data to investigate the contributions to the price discovery process by individual banks in the foreign exchange market.

- "Intraday Lead-Lag Relationships Between the Futures, Options and Stock Market", F. de Jong and M.W.M. Donders. Includes interesting method to estimate cross corelation whith asynchronous trades.

- Cointegration in Single Equations in lectures from Ronald Bewley

- Vector Error Correction Modeling from SAS online support.

- Variance ration testing: an other test for stationary aplication to stokc market indices, cached random walk or mean reversion ..., cached On the Asymptotic Power of the Variance Ratio Test, cached

- a general discussion on Econometric Forecasting and methods, by P. Geoffrey Allen and Robert Fildes.

- a simple presentation of Dickey Fuler test in French, cached

- The R tseries package include Augmented DickeyFuller

- How to do a 'Regular' Dickey-Fuller Test Using Excel cached

bibliography list from Petits Déjeuners de la Finance

- Alexander, C. (1994): History Debunked RISK 7 no.12 (1994) pp59-63

- Alexander, C. (1995): Cofeatures in international bond and equity markets Mimeo

- Alexander, C., Johnson, A. (1992): Are foreign exchange markets really efficient ? Economics Letters 40 (1992) 449-453

- Alexander, C., Johnson, A. (1994): Dynamic Links RISK 7:2 pp56-61

- Alexander, C., Thillainathan, R (1996): the Asian Connections Emerging Markets Investor 2:6 pp42-47

- Beck, S.E.(1994): Cointegration and market inefficiency in commodities futures markets Applied Economics 26:3 pp 249-57

- Bradley, M., Lumpkin, S. (1992): The Treasury yield curve as a cointegrated system Journal of Financial and Quantitative Analysis 27 pp 449-63

- Brenner, R.J., Kroner, K.F. (1995): Arbitrage, cointegration and testing the unbiasedness hypothesis in financial markets Journal of Financial and Quantitative Analysis 30:1 pp23-42

- Brenner, R.J., Harjes, Kroner, K.F. (1996): Another look at alternative models of the short term interest rate Journal of Financial and Quantitative Analysis 31:1 pp85-108

- Booth, G., Tse, Y. (1995): Long Memory in Interest Rate Futures Markets: A Fractional Cointegration Analysis Journal of Futures Markets 15:5

- Campbell, J.Y., Lo, A.W., MacKinley, A.C. (1997): The Econometrics of Financial Markets Princeton University Press

- Cerchi, M., Havenner, A. (1998): Cointegration and stock prices Journal of Economic Dynamic and Control 12 pp333-4

- Chowdhury, A.R. (1991): Futures market efficiency: evidence from cointegration tests The Journal of Futures Markets 11:5 pp577-89

- Chol, l. (1992): Durbin-Haussmann tests for a unit root Oxford Bulletin of Economics and Statistics 54:3 pp289-304

- Clare, A.D., Maras, M., Thomas, S.H. (1995): The integration and efficiency of international bond markets Journal of Business Finance and Accounting 22:2 pp313-22

- Cochrane, J.H. (1991): A critique of the application of unit root tests Jour. Econ. Dynamics and Control 15 pp275-84

- Dickey, D.A., Fuller, W.A. (1979): Distribution of the estimates for autoregressive time series with a unit root Journal of the American Statistical Association 74 pp427-9

- Duan, J.C., Pliska, S. (1998): Option valuation with cointegrated asset prices Mimeo

- Dwyer, G.P., Wallace, M.S. (1992): Cointegration and market efficiency Journal of international Money and Finance

- Engle, R.F., Granger, C.W.J. (1987): Cointegration and error correction: representation, estimation and testing Econometrica 55:2 pp251-76

- Engle, R.F., Yoo, B.S. (1987): Forecasting and testing in cointegrated systems Jour. Econometrics 35 pp143-59

- Granger, C.W.J. (1988): Some recent developments on a concept of causality Jour. Econometrics 39 pp199-211

- Hail, A.D., Anderson, H.M., Granger C.W.J. (1992): A cointegration analysis of Treasury bill yields The Review of Economics and Statistics pp116-26

- Hamilton, J.D. (1994): Time Series Analysis Princeton University Press

- Harris, F.deB., McInish, T.H., Shoesmith, G.L., Wood, R.A. (1995): Cointegration, Error Correction, And Price Discovery On Informationally Linked Security Markets Journal of Financial and Quantitative Analysis 30:4

- Hendry, D.F. (1986): Econometrics modelling with cointegrated variables: an overview Oxford Bulletin of Economics and Statistics 48:3 pp201-12

- Hendry, D.F. (1995): Dynamic Econometrics Oxford University Press

- Johansen, S. (1988): Statistical analysis of cointegration vectors Journal of Economic Dynamics and Control 12 pp231-54

- Johansen, S., Juselius, K. (1990): Maximum likelihood estimation and inference on cointegration - with applications to the demand for money Oxford Bulletin of Economics and Statistics 52:2 pp169-210

- Masih, R. (1997): Cointegration of markets since the '87 crash Quaterly Review of Economics and Finance 37:4

- Proietti, T. (1997): Short-run dynamics in cointegrated systems Oxford Bulletin of Economics and Statistics 59:3

- Schwartz, T.V., Szakmary, A.C. (1994): Price discovery in petroleum markets: arbitrage, cointegration and the time interval of analysis Journal of Futures Markets 14:2 pp147-167

- Schmidt, P., Phillips, P.C.B. (1992): LM tests for a unit root in the presence of deterministic trends Oxford Bulletin of Economics and Statistics 54:3 pp257-288

- Wang, G.H.K., Yau, J. (1994): A Time Series Approach To Testing For Market Linkage: Unit Root And Cointegration Tests Journal of Futures Markets 14:4

11 References

[1] Alexander, C and Dimitriu, A. *"Cointegration-based trading strategies: A new approach to enhanced index tracking and statistical arbitrage"* , 2002. cached. Discussion Paper 2002-08, ISMA Centre Discussion Papers in Finance Series. 26

[2] Alexander, C, Giblin, I, and Weddington, W. *"Cointegration and asset allocation: A new active hedge fund strategy"* , 2002. cached. Discussion Paper 2003-08, ISMA Centre Discussion Papers in Finance Series. 2, 26

[3] Chambers, M. J. *"Cointegration and Sampling Frequency"* , 2002. cached. 26

[4] Cont, R. *"Empirical properties of asset returns - stylized facts and statistical issues"* . QUANTITATIVE FINANCE, 2000. cached. 25

[5] Doornik, J. A and O'Brien, R. *"Numerically Stable Cointegration Analysis"*, 2001. cached. 26

[6] Engle, R and Granger, C. *"Cointegration and Error-Correction: Representation, Estimation and Testing"* . Econometrica, 55:251–276, 1987. 26

[7] Goetzmann, W, g. Gatev, E, and Rouwenhorst, K. G. *"Pairs Trading: Performance of a Relative Value Arbitrage Rule"* , Nov 1998. cached. 10

[8] Hasbrouck, J. *"Intraday Price Formation in US Equity Index Markets"* , 2002. cached. 26

[9] Herlemont, D. *"Optimal Growth"* , 2003. cached. Discussion papers. 11

[10] Kargin, V. *"Optimal Convergence Trading"* , 2004. cached. 11

[11] Murray, M. P. *"A drunk and her dog : An illustration of cointegration and error correction"* . The American Statistician, Vol. 48, No. 1, February 1994, p. 37-39. cached. 2, 26

[12] Stock, J. H and Watson, M. *"Variable Trends in Economic Time Series"*. Journal of Economic Perspectives, 3(3):147–174, Summer 1988.

[13] Trapletti, A, Geyer, A, and Leisch, F. "*Forecasting exchange rates using cointegration models and intra-day data*" . Journal of Forecasting, 21:151–166, 2002. cached. 26

[14] Thompson, G. W. P. "*Optimal trading of an asset driven by a hidden Markov process in the presence of fixed transaction costs*" , 2002. cached. 11