Learning and Trusting Cointegration in Statistical Arbitrage

by Richard Diamond*

A peer-reviewed version of this article has been published as a technical paper in *WILMOTT Magazine*, November 2013, Issue 68.[1]

Abstract

The paper offers adaptation of cointegration analysis for statistical arbitrage. Cointegration is a structural relationship model that relies on dynamic correction towards the equilibrium. The model is ultimately linear: when relationships are decoupled, the forecast of individual price follows a linear trend over the long term. 'Error correction' terminology does not apply to forecasting and therefore, is misleading.

Dynamic correction towards the equilibrium realises as a mean-reverting feature of the spread generated by the cointegrated relationship. Quality of mean-reversion defines suitability for statistical arbitrage and is evaluated by fitting to the Ornstein-Uhlenbeck process. Trade design cannot rely on standard cointegration tests due to their low power; their formulation is incompatible with the GBM process for asset price. However, the econometric specification of equilibrium correction is compatible with the OU process fit.

There are two technical appendices. Appendix A collects time series decompositions and derivations frequently omitted in presentation of the equilibrium correction. Appendix B discusses the common issues of equity pairs trading that relies on simple cointegration.

Keywords: time series decomposition, forecasting, cointegration, equilibrium correction, mean-reversion, Ornstein-Uhlenbeck, spread trading, statistical arbitrage

JEL Classification: C5, C62, G10

1 Linear Simplicity of Cointegration Model

Cointegration is known as a model for a long-term relationship, which is a shortcut taken for the ease of explanation. It is illustrative to contrast cointegration and correlation while, in fact, multiple scenarios are possible: correlated series can run away from each other, while cointegrated series might be seemingly unrelated. Correlation suffers

*Fitch Learning, 4 Chiswell Street, London EC1Y 4UP, UK. *E-mail: r.diamond@cqf.com*

[1] This Version: September 30, 2014 offers a clarification of the OU process fitting in Appendix A

from shortcomings of its measurement: it reflects co-movement on the small timescale only and, when measured over the long term, compounds to abnormally high values. Correlation among equities is usually range-bound, so it is a fundamentally uncertain parameter. Cointegration too have an element of uncertainty: **the spread produced by cointegrated time series is stationary and mean-reverting *around* some linear equilibrium level**. The tie-in among cointegrated series is not exact.

When learners are introduced to cointegration analysis, they are guided through a number of *supporting* concepts of vector autoregression, stationarity, how time series respond to shocks, how time series can be decomposed, and how we go about testing for that unit root that makes series a non-stationary accumulator of shocks. While it is necessary to introduce these concepts as building blocks of cointegration analysis, the prolonged nature of introduction is an obstacle for learning. In this note, I guide through 'model specification' for cointegration analysis while collecting the requisite transformations and hints in ways accessible to the learner. I show how cointegration can be an intuitive and attractive concept for trading a basket of instruments (spread).

It helps an introductory explanation to list the uses of cointegration at first sight: pairs and groups trading in equities and plain spread (basis) trading in fixed income. Less exploited strategies are designed using several instruments across the term structure (yield curve) which requires the use of multivariate cointegration. The desired stationary spread is often hidden behind the noise generated by stochastic (integrated) processes and requires the analysis of data projections in orthogonal dimensions. Filtering out a mean-reverting and tradable spread is the essence of statistical arbitrage.

Econometric models, such as cointegration, were built on what really is very low-frequency data (e.g., quarterly inflation rates and economic indicators). By comparison, the higher frequency asset price data (daily and above) is nothing but the noise. Cointegration analysis relies on representation of series as a sum of shocks and utilises the fact that *two or more time series have similar noise while growing at stable rates.* Then it becomes a question of combining non-stationary series in such a way that eliminates the common noise and produces the stationary spread ('a residual' of cointegrated relation):

$$\boldsymbol{\beta'_{Coint}Y_t} = \beta_1 y_{1t} + \beta_2 y_{2t} + \cdots + \beta_n y_{nt} \quad \sim \quad I(0) \tag{1.1}$$

Common regression notation is a foe here. The weights or hedging ratios are not the usual regression coefficients but restrictions estimated in a special way. Software packages normalise the cointegrating relationship to $\boldsymbol{\beta'}_{Coint} = [1, -\tilde{\beta}_2, \dots, -\tilde{\beta}_n]$—in this way, cointegration looks like differencing. Granger-Johansen representation (A.6) decomposes the dynamics between two or more non-stationary variables into two components:

1. Stochastic process in common that gets removed by that special 'differencing'.
2. Stationary spread that mean-reverts around the long-run equilibrium level.

The spread is a stationary and autoregressive process, which gives it the mean-reversion property. But it is not an integrated process, 'integrated of order zero'. In comparison, a non-stationary $y_{i,t}$ is 'integrated' because it can be decomposed into the infinite sum of residuals that are econometric *moving averages*.[2]

Building models of non-stationary time series by means of a regression—we will see below how regression for y_t is a static model —requires a genuine long-run relationship among the series. **In the absence of a long-run run relationship, there is no way to tie the non-stationary variables to produce a stationary spread.** A spurious regression will show the high goodness of fit R^2 but it has been proven that its coefficients are random variables that do not converge to their true value of zero. Therefore, the common statistical inference fails in the special case of spurious regression. However, it is not possible to detect that without building a further model for cointegration.

2 Learning Cointegration

2.1 Model Review

It has been my observation that design simplicity of cointegration model escapes the first-time learners. It is common that an analytical structure appears clear and simple to experts while being complex to novices who do not recognise the links. Let us see how we can start with a simple idea of the linear equilibrium and arrive at the multivariate model that accommodates the common stochastic trend as well as deterministic trends (i.e., growth with time or at a constant rate).

The familiar linear regression is **the** equilibrium model, suitable to model stationary variables only. If we regress two non-stationary series y_t, x_t, the stochastic trend in each will not allow to obtain a good model.

$$y_t = a + bx_t \tag{2.1}$$

We proceed to use the regression to model what we can, the growth rate Δy_t. How we moved from a regression for y_t to this fully equivalent regression for Δy_t is explained by derivations (A.3) to (A.4) in Appendix A.

$$\Delta y_t = \beta_g \Delta x_t - (1-\alpha)\left[y_{t-1} - a_e - b_e x_{t-1}\right] \tag{2.2}$$

$$\Delta y_t = \beta_g \Delta x_t - (1-\alpha) e_{t-1} \tag{2.3}$$

[2] MA(∞) decomposition (A.2) is a theoretical exercise but it reveals the stochastic trend in a non-stationary time series. It also completes A Quest for Invariance: asymptotically, any noise is an independently and identically distributed process (in its increments), *an iid invariant* of a model, Once such an iid invariant identified, it can be substituted with random numbers to generate projections, stress-tested, and, in the case of cointegraiton analysis, removed by the special differencing $\boldsymbol{\beta}'_{\boldsymbol{Coint}}\boldsymbol{Y_t}$.

The stationary spread sought is e_{t-1} but its role within the model is subtle: if cointegration exists the term is significant but its expectation of the equilibrium-correction term is equal to zero $\mathbb{E}[y_{t-1} - a_e - b_e x_{t-1}] = 0$. A financial quant would recognise this as the steady state model (we have a stationary spread series with a mean). The speed of correction towards the equilibrium level $-(1-\alpha)$ is an inferred (calibrated) parameter.

The model for Δy_t is known under two names: an equilibrium-correction mechanism and 'Error Correction Model'. The latter name is particularly misleading because the model does not correct any forecasting errors. Instead, it provides an adjustment towards the assumed long-run equilibrium. Next, I will give an overview the difficulties with the model's specification, use and understanding as more theory elements being plugged in.

2.2 Laborious Model Specification

Issue 1 Model specification for the multivariate autoregression, leading to cointegrated regression models, is daunting: it requires testing for significance of coefficients, autocorrelation, optimal lag, stability, weak stationarity, causality, time trend and exogenous variables. Model selection relies on the likelihood ratio test.

Being introduced to cointegration via theoretical and econometric route does not make for an easy journey of learning. Textbooks are written by experts whose understanding was transformed a long time ago and thus, they assume much about the readiness of readers. The aim provide an explanation competes with the aim to condense as much information as possible. In the end, the latter aim wins: time series transformations become omitted, notation under-explained, and model explanations overconnected. In order to acquire *a threshold concept* that cointegration is, one needs the actionable knowledge of (a) vector autoregression and (b) hypothesis testing – one's representation should operate faster than textbook chapters go. To achieve that, several reviews of the same material using different sources are always necessary.

It is my diagnosis that the inference of multivariate cointegration is under-explained in common time series textbooks. Therefore, I identified a set of four references (different sources) for learning cointegration: they are practical manuals that blend econometric theory with data analysis. If one needs a quick computational introduction, [2] is the sole practical guide with code to estimate cointegrating relationship matrix, in which maximum likelihood derivations given without creating several chapters. In this spirit, I present relevant time series decompositions and building blocks of the equilibrium-correction and cointegration models in Appendix A.

Issue 2 Calibration and uncertain parameters. Cointegrating relationships matrix $\boldsymbol{\Pi} = \boldsymbol{\alpha\beta'}$ is calculated **as if** it is factorised into cointegrating vectors $\boldsymbol{\beta}$ and the speeds of adjustment to the equilibrium $\boldsymbol{\alpha}$. The factorisation is not unique and require normalisation: $\boldsymbol{\beta}_{MLE,C}$ opens the space spanned by many cointegrating relations; without 'identifying restrictions' there could be no way of making sense about differencing weights/allocations/hedging ratios provided by cointegrating vectors. But somehow, one have to know the sensible restrictions on what otherwise are uncertain parameters.

Representing the cointegration estimation process schematically helps to see the unspoken assumptions. Estimating $\boldsymbol{\beta}$ *first* presupposes the potential long-run relationship before the model's significance is confirmed. It is possible to impose $\boldsymbol{\beta}$ as known rather than estimating from the data. Then, **the speed of adjustment becomes a calibrated parameter** fitted to the assumed long-run relationship.

$$\text{Estimate } \boldsymbol{\beta}_{MLE} \quad \Rightarrow \quad \text{Normalise } \boldsymbol{\beta}_{MLE,C} \quad \Rightarrow \quad \text{Calibrate } \boldsymbol{\alpha}$$

$$\Rightarrow \quad \text{Reconstruct} \qquad \boldsymbol{\Pi_{MLE}} = \boldsymbol{\alpha\beta'_{MLE,C}}$$

The entire estimation approach is 'a reverse engineering' that is not intuitive, but as often happens, mimics the evolution of thought preceding the model's invention: Glive Granger set out to prove that linear combinations of non-stationary variables remain non-stationary, however in the process, he identified the conditions under which the stationary spread could be observed [7]. This argument and its presentation earned the Nobel Prize for the advance in econometric analysis of non-stationary processes.

The key technical (matrix algebra) condition is for the relationships matrix $\boldsymbol{\Pi}_{n\times n}$ to have a reduced rank of $r < n$ linearly independent rows. Otherwise in model (A.8) (expressed in matrix form), the stationary $\Delta\mathbf{Y_t}$ would be equal to non-stationary $\boldsymbol{\Pi Y_{t-1}}$. If the condition holds, then it is a question of further matrix algebra to apply a rank-revealing decomposition and identify r cointegrating relations.[3] Financial quants would recognise the non-uniqueness of factorisation $\boldsymbol{\alpha\beta'}$ that stems from the reduced rank as a case of uncertain parameters—exact values unknown but their range or space is defined.

Issue 3 There could be deterministic trends in data that do not bundle together with cointegration. Matching the GBM and ECM is already problematic on equation level:

$$d\boldsymbol{Y_{t,\tau}} = \underline{\boldsymbol{Y_t\mu}dt} + \boldsymbol{Y_t\sigma}d\boldsymbol{W}_t \quad \Rightarrow \quad \Delta\mathbf{Y_t} = \boldsymbol{\Gamma}\Delta\mathbf{Y_{t-1}} + \underline{\boldsymbol{\alpha\beta'Y_{t-1}}} + \underline{\boldsymbol{\mu}_1 t} + \underline{\boldsymbol{\mu}_{0,\tau}} + \boldsymbol{\epsilon}_{t,\tau} \quad (2.4)$$

[3]For example, Gaussian elimination (LU decomposition) will show that any linearly dependent rows zero out revealing the independent rows and their number, the rank. Other decomposition methods are possible, and one can take the opportunity to revisit matrix algebra and represent matrix calculations in an explicit form as done in (A.9).

This blend of cointegration with other trends is an example of the most complex model specification of the Johansen Framework for multivariate cointegration analysis (A.8). In addition to the equilibrium-correction term it incorporates growth with time and constant growth trends (in (2.4) all three terms go in that order and are underlined).

Model-wise, the trends can be folded into the cointegrating relationship by looking for $\boldsymbol{\mu_1 t + \mu_0 = \alpha\rho_1 t + \alpha\rho_0}$ which will make VECM a restricted model. The question of how to model the trends (whether to restrict or not) is the central issue of the Johansen Framework – five model specifications are commonly considered in good sources [1][2][3].

A quick practical check for trends other than cointegration being present is to verify if $\mathbb{E}[\Delta \boldsymbol{Y}_t] = 0$ holds empirically. If the mean is not close to zero, constant and/or time-dependent growth trend is present. This works because $\mathbb{E}[\boldsymbol{\beta}'_{MLE,C}\boldsymbol{Y}_t] = 0$ or constant.

Monte-Carlo Experiment In order to check a number of assumptions about how trends can affect the specification of cointegration model, I run cointegration analysis on 20,000 simulated asset price paths that followed GBM with a range of drift values from 0.02 to 0.40. Matrices of observed frequency of cointegrating relationships were compiled for model specifications with linear growth trend, time-dependent trend (growth with time), and both. The higher frequency was observed for specifications in which cointegration and time-dependent trend were present but separated; in those models, coefficient $\boldsymbol{\mu}_1$ was extremely small, *less than other coefficients by* $O(10^{-2})$. Specifications with time-dependent trend folded into cointegrated relationship $\boldsymbol{\mu}_1 t = \boldsymbol{\alpha\rho_1} t$ identified the least number of such relationships.[4] This is good news because, de-trending a growing spread series (to trade the mean-reversion) would be difficult otherwise.

For the GBM model, it is expected that the drift value discretised over a daily time step will be small. A not so obvious question is how the long-term growth is achieved by a GBM path? The answer is: by adding the small positive value $\boldsymbol{Y_t\mu}dt$ to the shock $\boldsymbol{Y_t\sigma}d\boldsymbol{W}_t$. At any particular time step the shock term is likely to be much larger than drift term. In the model for $\Delta\boldsymbol{Y_t}$, the base GBM drift $\boldsymbol{\mu}dt$ corresponds to the constant growth rate $\boldsymbol{\mu}_{0,\tau}$.[5] The drift is likely to interfere with the stationarity of the spread because it has to be folded in cointegrating relationship for the VECM discretisation to match the GBM model. The restriction comes from a small-time, no jumps approximation as follows:

$$\frac{Y_t - Y_{t-\tau}}{Y_{t-\tau}} \approx \mu\tau \quad \Rightarrow \quad \boldsymbol{\Delta Y_t} = \boldsymbol{\mu}\tau\boldsymbol{Y_{t-1}} + \cdots \tag{2.5}$$

$$\boldsymbol{\Delta Y_t} = \boldsymbol{\mu}\tau\boldsymbol{Y}_{t-1} + \boldsymbol{\alpha\beta'Y_{t-1}} + \cdots \quad = \quad \boldsymbol{\alpha\rho_{0,\tau}Y_{t-1}} + \boldsymbol{\alpha\beta'Y_{t-1}} + \cdots \tag{2.6}$$

$$\mathbb{E}[\boldsymbol{\beta'Y_{t-1}}] = \boldsymbol{\mu}_e \quad \Rightarrow \quad \mathbb{E}[(\boldsymbol{\rho_{0,\tau}} + \boldsymbol{\beta'})\boldsymbol{Y_{t-1}}] = \boldsymbol{\mu}_e \tag{2.7}$$

[4] Model H^* of MATLAB *jcitest* function that implements the Johansen Framework. Model choice is to restricted trends to cointegrating relationship or add as the extra terms as in (2.4) model for $\Delta\mathbf{Y_t}$.

[5] The expected growth $\mathbb{E}[Y_t|Y_0] = Y_0 e^{\mu_{0,\tau}t}$ should not be confused with $\mu_1 t$ term.

Estimation $\boldsymbol{\rho_{0,\tau}}$ is subject to the effectiveness of the earlier calibration of $\boldsymbol{\alpha}$. The restriction of cointegration model (2.7) means adding to each of the cointegrating weights $\boldsymbol{\beta}'$ a quantity that supposed scale with time continuously (i.e., be a constant for each length of time period). Non-constance of the quantity $\boldsymbol{\rho_{0,\tau}}$ will make econometric tests to fail in detecting cointegration among GBM series. Non-robust (changing) estimates for the cointegrating weights $\boldsymbol{\beta}'$ and equilibrium level $\boldsymbol{\mu}_e$ pose a challenge for trade design.

We considered the underlined terms in (2.4) but there are effects coming from $\boldsymbol{\Gamma}$ and $\boldsymbol{\epsilon}_{t,\tau}$ terms. If each asset grows at a stable rate, then the relationship between growth rates (drifts) would also be stable and efficiently captured by the matrix $\boldsymbol{\Gamma}$ with robust estimates of growth ratios as matrix elements. In that case, the reliance of the model on the underlined drift terms (i.e., to achieve a good fit) would be reduced. Adding deterministic variables to the discretisation of a stochastic process is not a recipe for a flexible model and out of sample usefulness of that model.

Scaling of the shocks $\boldsymbol{\epsilon}_{t,\tau}$ with the asset price level in GBM means that they are modelled as extremely large compared to the equilibrium-correction term $\boldsymbol{\alpha\beta' Y_{t-1}}$. If an integrated process is a sum of its shocks then,

$$\sum \epsilon_t \quad \Rightarrow \quad \sum dW_t \quad \Rightarrow \qquad \boldsymbol{\epsilon}_{t,\tau} \equiv \int_t^{t+\tau} \boldsymbol{Y_t \sigma} dW_s \quad \sim N(0, \boldsymbol{\Sigma}_\tau)$$

Looking for the less disturbing shocks, the modellers employed the square root process (CIR) with $\boldsymbol{Y_t} \to \sqrt{\boldsymbol{Y_t}}$ and the Ornstein-Uhlenbeck process with diffusion not dependent on the asset level.

A conclusion that comes to fore: it is problematic to match the drift and diffusion terms of GBM SDE with the ones of VECM model developed in econometrics. I suggest **to see the GBM and VECM as alternative factorisations**. If we believe that both, growth according to the GBM and cointegration over the long run, are present in the data then two phenomena are unlikely to be conductive to each other.

2.3 Limited Forecasting

Issue 4 The very design of cointegration analysis limits its usefulness in forecasting: **cointegration is a structural model** that studies a set of long-run relationships within a multivariate system. Cointegration tests were developed as tests for conditions under which it is reasonable to assume that stochastically trending series are linked by the long-run equilibrium. The purpose of testing is detecting a spurious regression. Exogeneity of individual variables and forecasting were never the promises of model fathers.

Let's look into how a forecast can be constructed. The finding that $\Delta\mathbf{Y_t}$ might have a statistically significant equilibrium-correction term is insightful. However, in order to construct a forecasting equation for $\mathbf{Y_t}$, one need to undo the differencing that is by definition an information-loosing transformation. That is possible only under the assumption that leading variables are not subject to the equilibrium correction themselves, the 'weak exogeneity' assumption. The inevitable introduction of causality of $y_{1,t}$ depends on $y_{2,t}, \ldots, y_{n,t}$ kind comes at a price of making the forecast conditional on how well-known the behaviour of the independent variables is.

The model is limited to apply over the long run – the time period over which the correction towards the equilibrium realises can range from months for equities to years for interest rates. The common choice of data for cointegration analysis is **the low-frequency series (monthly or quarterly)**. It is an empirical fact that the short-run impact of the equilibrium correction is very small ($\boldsymbol{\alpha} \ll 0.1$) and addition of the extra terms $\Delta\mathbf{Y_{t-k}}$ or deterministic trends to the model for $\Delta\mathbf{Y_t}$ 'erases' the impact of the equilibrium correction term $\boldsymbol{\beta'Y_{t-1}}$. In a model with the constant growth, the short term forecast quickly converges to that constant $\boldsymbol{\mu}_{0,\tau}$.

$$\Delta\mathbf{Y_t} = \boldsymbol{\mu}_{0,\tau} + \boldsymbol{\alpha}\,[\boldsymbol{\beta'Y_{t-1}} - \boldsymbol{\mu_e}] + \boldsymbol{\epsilon}_{t,\tau} \qquad \Rightarrow \qquad \mathbb{E}[\boldsymbol{\Delta}\mathbf{Y_t}] = \boldsymbol{\mu}_{0,\tau} \tag{2.8}$$

If the equilibrium level shifts $\boldsymbol{\mu}_e \rightarrow \boldsymbol{\mu}_e^*$, then any model that relies on the equilibrium-correction will provide **a forecast in the opposite direction**. Given that non-stationary time series are prone to shifts from the shocks, the opposite-direction forecast is a common issue generating mistrust in cointegration models.

It is not uncommon to see recommendations to work with stationary data, for example, apply vector autoregression (VAR) to forecast returns and improve such forecasts by adding the equilibrium correction term because fundamentally, returns should be driven by the same market risk factors (APT). But after all 'improvements' to the model specification, one can be in for a surprise of the short-term forecasting error can be anything *O(200%)*. The correction term will add up to a noticeable trend over time, but the forecasting requires lower frequency than the data. Because of the loss of information from differencing, returns forecasting does not solve the problem of asset price prediction and is in contradiction with the idea of modelling returns as *iid* invariants. In efficient markets, it must not feasible to extract information from returns.

To arrive at a forecasting equation for $\widehat{\boldsymbol{Y}_t}$, one has to go through the labour of econometric model specification for the autoregression of non-stationary time series and, almost inevitably, be set up for disappointment about the usefulness of hard-earned results. First, the forecasting is based on the chain rule linking a forecast for time $T+h$

to the last observed value at time T:

$$\widehat{\boldsymbol{Y}}_{\boldsymbol{T+h}} = \widehat{\boldsymbol{\Pi}}^{\boldsymbol{h}} \boldsymbol{Y}_{\boldsymbol{T}} + \sum_{\boldsymbol{i=0}}^{\boldsymbol{h-1}} \widehat{\boldsymbol{\Pi}}^{\boldsymbol{i}} \boldsymbol{C}$$

where $\widehat{\boldsymbol{\Pi}}$ is **not** limited to cointegrating relationships $\boldsymbol{\alpha\beta'_{MLE,C}}$ but estimated by the generic vector autogregression. The powering up of the relationship matrix $\widehat{\boldsymbol{\Pi}}^{\boldsymbol{h}}$ on each time step h multiplies the uncertainty of the forecast [1]. The farther the projection, the wider its confidence interval bounds. Second, the outcome of long-term forecasting is trivial and known: growth at a constant rate, that can be seen as linear and time-dependent. But a trend-following strategy based on such forecast is extremely simplistic and prone to shocks. Asymptotic analysis of distributions produced by non-stationary time series suggest that impact of shocks is accumulated. Knowing that, **an experienced arbitrageur would not use the chain forecasting to construct a trade.** Instead, she would **trade the mean-reversion** of the spread generated by a cointegrated relationship.

3 Constructing A Spread Trade

It is common that the use and usefulness of a model depends on which of the two sides is using it. One must realise the difference between **the traders** who are speculators and arbitrageurs that can be neutral to asset price direction and **the hedgers** who are bound to work off forecasts and carry an implied belief into econometric models. A market-marking operation can assume both sides: while keeping only residual up or down exposure, it is interested in the directional forecast in order to seek cheaper hedges.

Cointegration for Hedgers The hedgers and forecasters desire for the model parameters to be stable and predictions to be reliable. They might employ regression models with time-varying parameters, fractional models, and variance reduction techniques in order to reduce the forecasting error.[6] The focus shifts to the elaborate model specification and efforts depend on stationarity and non-heteroskedastic variance (i.e., require some constance in variance and therefore, work better in low-volatility regimes). In the world of statistical arbitrage, the hedgers are at the mercy of their models' ability to capture the features of time series.

If used for the short-term hedging, the equilibrium-correction is likely to generate a regular loss because the more a growth rate Δy is different from its expectation, the

[6]The next steps are heterogeneous regression with time-varying parameters and using the space-state representation techniques, such as Kalman filtering. 'Time-varying' term is misleading because the parameters are really state-varying. Because it is the equilibrium level $\boldsymbol{\mu_e}$ that shifts, it makes sense to talk about 'conditional equilibrium' cointegration.

more severe correction is expected. That might work well for the low-frequency data but it is unrealistic to expect the complete after-shock correction for daily or higher frequency data. A hedge can generate some profit, but the forecaster is not prepared to let it run because of the expectation of a fast correction to the equilibrium. For non-stationary series, the equilibrium itself might be shifting which makes a profit to look like an unexpected accident. There are problems with the hedgers' worldview.

Cointegration for Traders A trader should seek to utilise the qualities of a phenomenon that a model brings to light. The phenomenon is cointegration: **in depth, the common factor driving behaviour of two or more time series**. The mean-reverting quality of a cointegrated spread $\boldsymbol{e_t} = \boldsymbol{\beta' Y}$ was empirically observed for various assets. The elegance of the equilibrium correction is how this spread appears inside all regressions linking changes in variables of the system $\Delta \mathbf{Y_t}$ (see (A.4) in Appendix A). The spread is referred to as 'a disequilibrium error' and gets thrown out in forecasting exercises and operations done under expectation. However, the arbitrageur's interest is not necessarily an improvement of a forecast but seeking to detect a particularly large disequilibrium error while minimising the model error.

Cointegration model is *an affine filter* applied to identify a mean-reverting and stationary spread, if exists. Traders take a theory-free approach to cointegration, one kind is offered in [4]. The empirical attributions emerging from the uses of a model can be more useful to traders than its pre-specified theoretical purpose or prescribed applications. The most useful attribution question for linear cointegraiton is: what is the common factor that drives an evolution of cointegrated assets?

The points made here rely on the Financial Modellers Manifesto [9] which can be extended with one more item: traded models include an affine derivation of the no arbitrage condition(s). In practical sense, it makes the output numbers robust (the output can be high-frequency) and allows to control P&L against an explicit mathematical result. **The cointegrated relationship that delivers $\mathbb{E}[\boldsymbol{\beta' Y_t} - \boldsymbol{\mu_e}] = 0$ itself represents such no arbitrage condition.** To design a trade it is necessary to backtest its P&L and clarify magnitude the drawdowns.[7] For a mean-reversion arbitrage, one requires the following items of information:

1. **Weights** for a set of instruments $\boldsymbol{\beta'_{Coint}}$ to construct a spread.
2. **Speed of mean-reversion** in the spread θ to evaluate the feasibility of a trade.
3. **Entry and exit levels** together with the equilibrium level $\boldsymbol{\mu}_e$, around which mean-reversion occurs. Fitting a spread to the Ornstein-Uhlenbeck process allows calculating the bounds as $\sigma_{eq} \approx \sigma_{OU}/\sqrt{2\theta}$, the proof can be found in (A.20).

[7] Generally, the P&L distribution must be studied. Explicit mathematical results can extend as far as expected variation in the P&L.

The weights do not need to come from a refined, Johansen-estimated cointegration model. It is possible to check if there is a significant speed of adjustment towards the equilibrium $\boldsymbol{\alpha}$ for any set of weights. This problem of non-unique factorisation haunts a cointegrated system of variables and is the problem for hedgers. The traders should be interested in **the quality of mean-reversion**, a broad concept that covers the stationarity of the spread, robustness of estimated cointegrating weights as well as how far the disequilibrium error can go and how often it reduces back to zero at the mean level $\boldsymbol{\mu_e}$ (half-life).

Proof (A.25) shows that the spread $\boldsymbol{e_t} = \boldsymbol{\beta' Y_t}$ is an AR(1) process which makes it technically possible to perform **fitting to the Ornstein-Uhlenbeck process** and evaluate the speed of mean-reversion ('recall force') θ: the higher the θ the more stronger and less fractionated the cointegrated relationship is. That is possible to establish **without conducting formal econometric tests for cointegration**. Other fitted parameters, the equilibrium level μ_e and σ_{eq}, offer to calculate z-score given that $e_{t,\tau\to\infty} \sim N(\mu_e, \sigma_{eq}^2)$. OU fitting recipe uses the VAR(1) model and is surprisingly simple (A.12).

Comparison of the OU SDE solution to Equilibrium Correction mechanism (A.23) gives an insightful representation, where all parameters fall in correct places with identifiable roles. In particular, the speed of adjustment towards the equilibrium is proportional to the autocorrelation of the OU process $\alpha \propto e^{-\theta\tau}$, which is mean-reverting.

In order to give an indication of time required to generate the $P\&L$, the reversion speed can be presented as **a global half-life**: the time between the equilibrium situations of spread e_t being at μ_e.

$$\widetilde{\tau} \propto \frac{\ln 2}{\theta}$$

Mean-reversion begins with $\theta > 0$. The higher the theta the shorter half-life of reversion to the mean, and the better the quality of mean-reversion becomes. It is not uncommon to observe $\theta = 10 \dots 200$ for the higher-frequency data of instruments linked by a term structure (e.g., futures and interest rate swaps).

An example of the stationary and mean-reverting spread was obtained from a cointegrating relationship among four futures linked to a volatility index. The observation sample covered 142 trading days from March to September 2012. Figure 3.1 shows the spread $e_t \sim I(0)$ fitted to the Ornstein-Uhlenbeck process with μ_e mean and σ_{eq} bounds. Cointegration acts as a filter: *for daily data, the spread has a longer half-life.* Empirically for this example, the reversion to the equilibrium level μ_e occurs circa every 22 days.

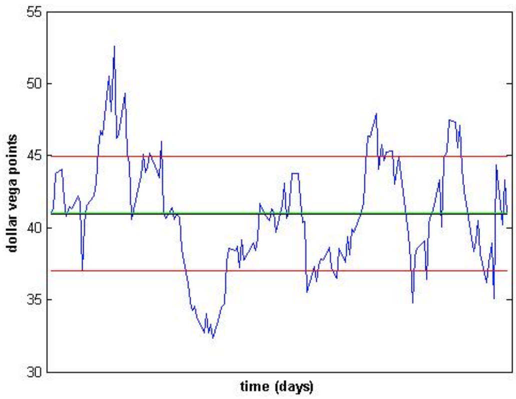

Figure 3.1: Mean-reversion in a cointegrated spread e_t, $\theta = 42.39$

'Trading half-life' is a default arbitrage strategy: it requires entering at a bound and exiting on the mean thus, generating half the possible alpha. Trading over horizons suggested by half-life might not be possible but there is an advantage in seeing the dynamics operating beyond the near-term horizons of other implied volatility traders.

The situation of near-cointegrated series is of consequence to trade design because it generates mean-reversion not suitable for trading the range of $\pm n\sigma$ above and below the equilibrium level. (1) The common situation is when the spread tends to be one-sided *wrt* to the mean. It indicates that the equilibrium level is prone to be shifting. Approaching the situation, there are two further distinctions: (2) the spread mean-reverts but trends (up or down), (3) the spread 'spends' some time around the mean. In both cases, it is still possible to observe the mean reversion *wrt* the original level but the full swings from upper to lower bound are rare. Rather than catching rare events, the arbitrageurs particularly ones working with high frequency, implement strategies that are based on crossing the mean. It is the safer choice when the spread is produced by near-cointegrated combinations. This is illustrated in Appendix B. In addition, having an attribution for the common stochastic process (that is typically a risk factor) allows to construct trades with conviction even if cointegration is fractionated and the spread is ill-behaving.

4 Trusting Cointegration

"In some analysis cointegrating combinations of two time series mean that stationarity [in spread] becomes a reasonable assumption." [1]

Cointegration model reveals a deterministic trend of specific kind: equilibrium-correction over the long run. Time series decompositions (A.2) and (A.6) as well as models (A.4) and (A.11) demonstrate that it is possible to remove the stochastic trend from an integrated time series and construct a stationary as well as autoregressive spread.

The decompositions are generic and convincing. The attribution of the common stochastic process is usually traceable to the impact of a systemic risk factor. The result is achieved by a linear combination(s) $\boldsymbol{\beta}'_{\mathbf{Coint}}\mathbf{Y}$ that transforms several non-stationary variables into the stationary spread. This theory-free and computationally easy result of cointegration analysis is attractive to arbitrageurs. Any statistical arbitrage strategy relies on mean-reversion somewhere.

The use of cointegration analysis in trading is not an issue of blind trust into the econometric model. It is a question of conditions under which the model delivers mean-reversion. Let us formulate three frequent technical complaints and address them, while not necessarily disagreeing:

1. The simplicity of mean-reversion around a linear equilibrium level is not to be trusted. What is the nature of the equilibrium, economic or 'technical'? One have to find the half-life for the mean-reversion.

2. The spread is slow to converge: for instance, the half-life times (calculated using the Ornstein-Uhlenbeck process) for cointegrated equity combinations are too long. This is often the case for unfiltered data.

3. Cointegration tests, including ones adjusted for other trends and autocorrelation, tend to find cointegration when 'it is not present'. Significance testing for the multivariate cointegration (Johansen Framework) is sensitive to deterministic trends and how they are specified in the regression equation.

Addressing Complaint 3 High sensitivity and low power of cointegration tests (be it the Dickey-Fuller, Phillips-Perron, or multivariate Johansen) are a direct consequence of how stationarity tests are constructed: they are more likely to accept that the series is integrated than reject this hypothesis. Cointegration is reported when the series are, in fact, near-cointegrated, $\beta \approx 1$. However, having a clean case of $\beta = 1$ is important for elimination of the common stochastic process. On the other end of empirical evidence, there is an observation that cointegration tests can fail to detect the relationship, particularly if time series have different starting values. Failure of econometric tests to detect cointegration was observed even for the time series that were simulated as a pair (i.e., returns of one asset leading another) [6].

There is a usability problem with standard econometric tests for cointegration. The cointegration model uses the equilibrium correction to make regression analysis applicable to non-stationary series but fails to give information necessary to complete a trade design. Cointegration analysis alone is not a solution for automated model selection and flow trading. It should be used on series that are pre-filtered by some fundamental criteria or model and pre-processed to remove other trends (i.e., normalised to start at the same point). It follows that testing all combinations of marketable assets for cointegration is nether a computationally efficient nor reliable approach.

Addressing Complaint 2 Sometimes the complaint is talks about processes 'rooted' in the Ornstein-Uhlenbeck dynamics as if that were a true representation of the data. The solution to the Ornstein-Uhlenbeck SDE gives a powerful tool to evaluate the quality of mean-reversion. One just has to look at the value of θ without the need to set up an elaborate specification for cointegration testing. The analysis of how the Ornstein-Uhlenbeck process matches features of cointegrated spread $\boldsymbol{e_t}$ is presented in Appendix A.7. The fitting is surprisingly simple, by estimation of VAR(1) only. The match between the regression and O-U process parameters is affine, see (A.13) and (A.14). For cointegration to exist, the technical condition is that *some* eigenvalues of transition matrix $\boldsymbol{\Theta}$ have positive real parts [4]. If that fairly relaxed numerical condition is satisfied, the spread $\mathbf{e_t}$ does converge to its deterministic drift $\boldsymbol{\mu_e}$.

Fitting to the Ornstein-Uhlenbeck process need not be the final stage of 'suitability for trading' analysis. Financial time series carry a mix of phenomena: cointegration, compound autocorrelation, and long memory – all mediated by the heteroskedasticity of variance. It is possible to find all of these phenomena for interest rates instruments. In this case, the modelling has to go back to the first principles.

The Johansen Framework for cointegration offers a decomposition into one stochastic trend (in general, the Levy Process) and several deterministic trends. The avant-garde of econometric research took turn in a technical attempt to address the shortcomings of cointegration by **modelling the stochastic trend with the Fractional Brownian Motion**, a continuous-time model for the long memory [8].

The Fractional Brownian Motion has autocorrelation that decays according to the power law τ^{2d-1} which is slower than the exponential pattern of the Ornstein-Uhlenbeck $e^{-\theta\tau}$. The dual nature of the Fractional Brownian Motion allows to model integrated series I(1) by setting parameter $d \approx 0$ as well as stationary-like series with low values of the Hurst exponent $H < 0.2$.[8] Interest rate swaps with pronounced long memory are modelled with $H > 0.5$ as the change increments are not the *iid* invariants (think of a bond getting closer to maturity). Simulated fractionally cointegrated series exhibit different strength of cointegration at different time periods, which is close to how real financial series behave. The model also addresses the issue of state-dependent equilibrium.

[8]The Hurst exponent is defined as $H = d + \frac{1}{2}$. The exact values $H = \frac{1}{2}$, $d = 0$ recover the Standard Brownian Motion.

Addressing Complaint 1 While impact over the long run is the definition of equilibrium, whether or not the 'global' equilibrium applies to the specific data is a matter of belief. The business of econometric analysis is to make the equilibrium-based models work (i.e., reveal and stabilise the equilibrium) by the means of conditioning on any number of things: low frequency of observations, particular time periods, exogenous parameters. The method assumes that time series are integrated with a removable stochastic trend (model risk).

The theories of asset pricing postulate the long-term equilibrium (1) between equity price and expected dividend cashflow (Gordon growth model) as well as (2) among asset returns exposed to the same market risk factors (APT). Empirical investigations of cointegration should include such models explicitly, either as pre- or post filters on the data. The econometric tests of Johansen Framework for multivariate cointegration with trends are **unlikely** to uniquely identify these specific relationships even if they were used to generate data [6].

The presence of the equilibrium is clear in two somewhat opposite cases:

Case 1. Large samples of low frequency data collected over the long term.
Case 2. Large filtered samples of the very high frequency data.

In between, there are grey cases that challenge the econometric tests that ultimately rely on closed-form decompositions and equilibrium-correction term (A.4). The term is usually small but significant, which confirms at least the conditional equilibrium.[9] Its impact is mediated by the calibrated speed of adjustment parameter $\boldsymbol{\alpha}$.

The small impact of the equilibrium-correction that adds up over the long run can be easily disturbed with the short-term shocks. While the theory suggests that it would take a while to correct such shocks, **the market correction is commonly observed as the large shock in the opposite direction** (please see spread examples in Appendix B). **That is more consistent with no arbitrage conditions at work in financial markets** (i.e., stability of some relative pricing) **rather than the idea of a long-run equilibrium**. It also suggests that the strategy of trading half-life is likely to be sub-optimal.

The problem of equilibrium detection is directly related to non-constancy of the rate at which an asset appreciates. If the rate would be a constant scalable with time as assumed by the Geometric Brownian Motion, there would be no technical problem of

[9]Otherwise, the term would be insignificant. The specification of the equilibrium-correction term is unequivocal in that it represents the long run equilibrium.

identifying its impact on the cointegration model and selecting the appropriate deterministic trend to augment the tests. However, the evidence, including a brief Monte-Carlo experiment conducted for this study, suggests that the asset's growth and equilibrium correction interfere with each other to a degree that it is technically difficult to recover known parameters of simulated processes.

The insightful analysis by Attilio Meucci [4] of the geometry of multivariate cointegrated system, standardised *wrt* its transition speeds $\boldsymbol{\Theta}$, shows how the shift towards the cointegrated plane occurs along the principal eigendirection, while the noise movements in other dimensions zero out at an exponential rate. This finding is common to principal component analysis of several kinds of financial series (i.e., forward rates and futures). **The cointegrated spread $\boldsymbol{e_t}$ can be seen as an attractor in its own right**. Empirical observations of equity pairs show recurrently that even near-cointegrated spreads maintain a stationary level.

If **Case 1** low-frequency cointegration is an economic phenomenon that is hard to trade on, **Case 2 high-frequency cointegration analysis** can fall back to the generic no-arbitrage interpretation (*wrt* the order book). High frequency data reflects a deterministic world: even for a low-depth book, the price in the next millisecond is not likely to be *much different* from before but likely to be *somewhat different* because of the dynamic matching of supply and demand in the order book. The dynamics guarantee conditional autoregression and a possibility to fit the Ornstein-Uhlenbeck process to time series directly, over short time horizons. Prices themselves are close to being martingales. There is no drift to be concerned with, and no trends can be detected except the stochastic one. A theory-free cointegration filter removes such integrated stochastic trend of many near-*iid* movements[10] and delivers a robust spread that is still high-frequency. The model provides that the spread is more likely to remain stationary than the underlying time series.

Cointegration must be utilised in the spirit of data exploration: attention to model specification should not become a preoccupation with statistical testing. Fischer Black noted that significance tests themselves are almost of no value for "the real world of research", and the best approach would be "to explore a model" [10]. Obtaining applicable results from the cointegration does **a.** making and testing assumptions about deterministic trends and **b.** pre-processing of data by either projecting in orthogonal dimensions and operating with principal components or identifying cointegration candidates by fundamental criteria, such as factor models available for asset allocation.

[10]Larger jumps are a consequence of order toxicity which makes them previsible. They can also be modelled as structural breaks.

It is better using an affine model in order to make data alive rather than merely relying on specifications of statistical tests as coded into software packages. This is the argument behind using an affine Ornstein-Uhlenbeck process fit versus the VECM. The Johansen Framework might tell a story about the structure of the relationship matrix of a multivariate autoregressive model but it suffers from uncertain attribution as much as any formal decomposition method does. Treating cointegration analysis as a set of statistical tests that simply points out a robust arbitrage with high significance values is akin a medieval anatomy experiment: it would overturn the unrealistic beliefs but be no closer to reproducing an alive arbitrage outcome.

Richard V. Diamond helps to run the Certificate in Quantitative Finance where he is responsible for setting implementation projects and cross-asset quantitative analytics. Prior to joining the CQF Team at Fitch, Richard worked for a private investment office. He advises on private equity, investment performance and effectiveness of hedges. The focus of his ongoing modeling is term structure arbitrage and algorithmic trading.

References

[1] *Explaining Cointegration Analysis: Parts I and II*
by David Hendry and Katarina Juselius, 2000 and 2001. Energy Journal.
See also *Teaching Undergraduate Econometrics at Oxford* at
`https://www.youtube.com/watch?v=We12iDwwYEE`

[2] *User Guide for Johansens Method* by Kyungho Jang and Masao Ogaki, 2001.

[3] *Modeling Financial Time Series with S-PLUS* by Eric Zivot and Jiahui Wang, 2006. See specifically Chapter 12 on VAR Models and Cointegration

[4] *Review of Statistical Arbitrage, Cointegration, and Multivariate Ornstein-Uhlenbeck* by Attilio Meucci, 2010. Available at `http://www.symmys.com/node/132`

This set of four references is recommended to continue learning about cointegration: they are practical manuals that blend econometric theory with data analysis.

[5] *The Misleading Value of Measured Correlation* by Mahdavi Damghani et al., 2012. WILMOTT Magazine

[6] *Cointegration Based Statistical Arbitrage* by Markus Harlacher, 2012. Masters Thesis, Swiss Federal Institute of Technology Zurich (ETH Zurich).

[7] *Professor Sir Clive W.J. Granger and Cointegration* by David Hendry, 2010. Journal of Financial Econometrics, 2010, Vol. 8, No. 2, 162-168.

[8] *Likelihood inference for a fractionally cointegrated vector autoregressive model* by Soren Johansen and Morten Nielsen, Econometrica, 2012, 80, 2667-2732.

[9] *Frequently Asked Questions in Quantitative Finance* by Paul Wilmott, 2009.
The Financial Modelers' Manifesto is written together with Emanuel Derman.

[10] *On Fischer Black: Intuition is A Merging of the Understander with the Understood* by Emmanuel Derman, 2009.

A Appendix A

In this appendix, I summarise the useful transformations that are often implied, when vector autoregression is used in order to build further models.

A.1 Decomposing AR as MA Process

We start with the first trick, formally known as the Wold's theorem: any stationary process ARMA(p, q) can be expressed as ARMA$(0, \infty)$ using the expansion:

$$\frac{1}{1-\beta L} \approx 1 + \beta L + \beta^2 L^2 + \cdots \equiv \sum_{i=0}^{\infty} (\beta L)^i \qquad \text{for lag operator } L \tag{A.1}$$

For example, we can move from AR(1) process $y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$ to MA(∞) process by rearranging $(1 - \beta_1 L) y_t = \beta_0 + \epsilon_t$ and applying the expansion

$$y_t = \frac{\beta_0}{1-\beta_1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_1^2 \epsilon_{t-2} + \cdots \qquad \text{s.t. } |\beta_1| < 1 \text{ gives stationary } y_t \tag{A.2}$$

The result is more important for its theoretical insight because empirical estimation of the infinite set of parameters is not feasible. Think of this as an equivalent to Taylor series expansion in statistical world.

A.2 Dynamic Regression to Equilibrium Correction

The familiar linear regression is **the** equilibrium model. It is also a static model that assumes stationary y_t and x_t. A dynamic regression model uses several lagged values:

$$y_t = \alpha y_{t-1} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \epsilon_t \tag{A.3}$$

We would like to re-specify the model to equilibrium-correction form

$$\begin{aligned}
y_t - \underline{y_{t-1}} &= \alpha y_{t-1} - \underline{y_{t-1}} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} - \underline{\beta_1 x_{t-1}} + \underline{\beta_1 x_{t-1}} + \epsilon_t \\
\Delta y_t &= -(1-\alpha) y_{t-1} + \beta_0 + \beta_1 \Delta x_t + (\beta_1 + \beta_2) x_{t-1} + \epsilon_t \\
&= \beta_1 \Delta x_t - (1-\alpha) \left[y_{t-1} - \frac{\beta_0}{1-\alpha} - \frac{\beta_1 + \beta_2}{1-\alpha} x_{t-1} \right] + \epsilon_t \\
&= \beta_1 \Delta x_t - (1-\alpha) \left[y_{t-1} - a_e - b_e x_{t-1} \right] + \epsilon_t \qquad \text{(A.4)} \\
&= \beta_1 \Delta x_t - (1-\alpha)\, e_{t-1} + \epsilon_t
\end{aligned}$$

where $y_t = a_e + b_e x_t$ is a static equilibrium model encapsulated within the relationship between non-stationary (integrated) x_t and y_t.

The models (A.3) and (A.4) are fully equivalent.

A.3 Granger-Johansen Representation (Perfect Unit Root)

We can represent a perfect unit-root process $x_t = x_{t-1} + \epsilon_{x,t}$ as a stochastic process that allows different 'trend' at any point in time.

$$x_t = \underbrace{\sum_{s=1}^{t} \epsilon_{x,s}} + x_0 \tag{A.5}$$

Regressing another perfect unit-root process y_t on x_{t-1} and substituting

$$y_t = x_{t-1} + \epsilon_{y,t} = \underbrace{\sum_{s=1}^{t} \epsilon_{x,s}} - \epsilon_{x,t} + \epsilon_{y,t} + X_0 \tag{A.6}$$

We see how x_t and y_t have **a stochastic process in common**. In continuous time, summation becomes integration and two series are **cointegrated**.

Subtracting (A.5) from (A.6) eliminates the common stochastic process and leaves the stationary autoregressive residual $e_t = \epsilon_{y,t} - \epsilon_{x,t} \equiv (y_t - b_e x_t - a_e)$.

In the multivariate setting, the joint stationary residual is written as $\mathbf{e_t} = \boldsymbol{\beta}'_{\mathbf{Coint}} \mathbf{Y_t}$ and called a vector error term.

A.4 Vector Error Correction Model (Equilibrium Correction)

The model for $\Delta \mathbf{Y_t}$ includes a vector term $\boldsymbol{\alpha} \mathbf{e_{t-1}}$ that corresponds to equilibrium correction in bivariate case (A.4). Notice change of convention $-(1-\alpha) \to \alpha$.

$$\Delta \mathbf{Y_t} = \boldsymbol{\Gamma} \Delta \mathbf{Y_{t-1}} + \underline{\boldsymbol{\alpha} \mathbf{e_{t-1}}} + \boldsymbol{\epsilon}_t \tag{A.7}$$

In econometrics literature, the full VECM is often expressed as

$$\Delta \mathbf{Y_t} = \boldsymbol{\Gamma} \Delta \mathbf{Y_{t-1}} + \underline{\boldsymbol{\Pi} \mathbf{Y_{t-1}}} + \boldsymbol{\Psi} \mathbf{D_t} + \boldsymbol{\epsilon}_t \tag{A.8}$$

- $\boldsymbol{\Gamma}$ is a coefficient matrix for lagged differences $\Delta \mathbf{Y_{t-1}}$ (past growth rates).
- $\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'_{\mathbf{Coint}}$ is a factorised matrix. For the model to be consistent, $\boldsymbol{\Pi}$ must have a reduced rank, otherwise stationary $\Delta \mathbf{Y_t}$ would be equal to non-stationary $\mathbf{Y_{t-1}}$.
- $\boldsymbol{\Psi}$ is a coefficient matrix for exogenous variables $\mathbf{D_t}$.
  It is most useful as an indicator for regime-switching events (a dummy variable).
- $\boldsymbol{\epsilon}_t$ is an innovations process outside of cointegration that should be $I(0)$, have no autocorrelation (serial correlation) or other surprises that carry information.

A.5 Reduced Rank for A Relationship Matrix

Granger representation theorem states that if matrix $\mathbf{\Pi}$ has a reduced rank $r < n$ then it can be factorised as

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

$$(n \times n) = (n \times r) \times (r \times n)$$

1. Reduced rank means that matrix $\mathbf{\Pi}$ has only r linearly independent rows (columns), and therefore, the other $n - r$ rows can be obtained by linear combination.
2. r columns of $\boldsymbol{\beta}$ are cointegrating vectors, and $n-r$ columns are common stochastic trends (unit roots) of the system.
3. Adjustment coefficients $\boldsymbol{\alpha}$ show how fast equilibrium correction operates.

If we restrict $\mathbf{\Pi} \propto \boldsymbol{\beta}$ to r=2 cointegrating vectors, the result is in line with (A.4)

$$\boldsymbol{e}_t = \boldsymbol{\beta}'_{\mathbf{Coint}}\boldsymbol{Y_t} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1n} \\ \beta_{21} & \cdots & \beta_{2n} \end{bmatrix} \begin{bmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} \beta_{11}y_{1t} & +\cdots+ & \beta_{1n}y_{nt} \\ \beta_{21}y_{1t} & +\cdots+ & \beta_{2n}y_{nt} \end{bmatrix}$$

The vector error correction term for a whole system is

$$\boldsymbol{\alpha e_t} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \vdots & \vdots \\ \alpha_{n1} & \alpha_{n2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \beta_{1i}y_{it} \\ \sum_{i=1}^n \beta_{2i}y_{it} \end{bmatrix} = \begin{bmatrix} \alpha_{11}\sum_{i=1}^n \beta_{1i}y_{it} + \alpha_{12}\sum_{i=1}^n \beta_{2i}y_{it} \\ \vdots \\ \alpha_{n1}\sum_{i=1}^n \beta_{1i}y_{it} + \alpha_{n2}\sum_{i=1}^n \beta_{2i}y_{it} \end{bmatrix}$$

This is an example of $\boldsymbol{\alpha e_t} = \boldsymbol{\alpha\beta}'_{\mathbf{Coint}}\boldsymbol{Y_t} = \mathbf{\Pi}\boldsymbol{Y_t}$. The correction term for each equilibrium is present and reiterated for each horizontal regression of VAR system.

It makes sense to vectorise multiplication of coefficients $\boldsymbol{\alpha}, \boldsymbol{\beta}'$ when the relationship matrix $\mathbf{\Pi}$ presented alone. For 3×3 example with rank $r = 2$ and normalised weights,

$$\mathbf{\Pi} = \text{Vec}\left(\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix}\right) \text{Vec}\left(\begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} \\ 1 & -\beta_{22} & -\beta_{23} \end{bmatrix}\right) \tag{A.9}$$

where each Δy_{it} has its own speed of adjustment α_{i1} to the same equilibrium,

$$\begin{aligned} \Delta y_{1t} &= \alpha_{11}\left[y_{1t-1} - \beta_{12}y_{2t-1} - \beta_{13}y_{3t-1}\right] + \sum b_{1i}\Delta y_{it-1} + \epsilon_t \\ \Delta y_{2t} &= \alpha_{21}\left[y_{1t-1} - \beta_{12}y_{2t-1} - \beta_{13}y_{3t-1}\right] + \sum b_{2i}\Delta y_{it-1} + \epsilon_t \\ \Delta y_{3t} &= \alpha_{31}\left[y_{1t-1} - \beta_{12}y_{2t-1} - \beta_{13}y_{3t-1}\right] + \sum b_{3i}\Delta y_{it-1} + \epsilon_t \end{aligned}$$

Similar expressions can be constructed for another equilibrium $[1, -\beta_{22}, -\beta_{23}]$.

A.6 Ornstein-Uhlenbeck Process (Quality of Mean-Reversion)

We use the process in order to model **the residual of cointegrating relationship** (vector error term) $\boldsymbol{e_t} = \boldsymbol{\beta'_{Coint}}\boldsymbol{Y_t}$, given its statistical properties of being stationary AR(1) process with frequent mean reversion around $\mathbb{E}[\boldsymbol{\beta'_{Coint}}\boldsymbol{Y_t}] = \boldsymbol{\mu_e}$. Its SDE gives a continuous representation of an autoregressive process with the transition matrix $\boldsymbol{\Theta}$ and 'scatter generator' (dispersion) matrix $\boldsymbol{S}$ which, for the fully multivariate case, is

$$d\mathbf{e_t} = -\boldsymbol{\Theta}(\mathbf{e_t} - \boldsymbol{\mu_e})\,dt + \boldsymbol{S}\,d\mathbf{W_t} \tag{A.10}$$

The regression-like solution in continuous time $t + \tau$

$$\mathbf{e_{t+\tau}} = \left(\mathbf{I} - e^{-\boldsymbol{\Theta}\tau}\right)\boldsymbol{\mu_e} + e^{-\boldsymbol{\Theta}\tau}\mathbf{e_t} + \boldsymbol{\epsilon}_{t,\tau} \tag{A.11}$$

Vector Autoregression equivalent for a small time period τ is

$$\mathbf{e_{t+\tau}} = \mathbf{C} + \mathbf{B}\mathbf{e_t} + \boldsymbol{\epsilon}_{t,\tau} \tag{A.12}$$

Once VAR(1) is estimated, we can solve for square matrix $\boldsymbol{\Theta}$ and vector $\boldsymbol{\mu_e}$:

$$e^{-\boldsymbol{\Theta}\tau} = \mathbf{B} \quad \Rightarrow \quad \boldsymbol{\Theta} = -\frac{\ln \mathbf{B}}{\tau} \tag{A.13}$$

Coefficient $\mathbf{B}$ corresponds to the unconditional autocorrelation $\mathrm{Corr}[\mathbf{e_t}, \mathbf{e_{t+\tau}}] = e^{-\boldsymbol{\Theta}\tau}$.

$$\left(\mathbf{I} - e^{-\boldsymbol{\Theta}\tau}\right)\boldsymbol{\mu_e} = \mathbf{C} \quad \Rightarrow \quad \boldsymbol{\mu_e} = \frac{\mathbf{C}}{\mathbf{I} - \mathbf{B}} \tag{A.14}$$

If we use (A.2) to decompose VAR(1) process (A.12) into its MA(∞) representation

$$\mathbf{e_{t+\tau}} = \frac{\mathbf{C}}{\mathbf{I} - \mathbf{B}} + \boldsymbol{\epsilon}_{t,\tau} + \mathbf{B}\boldsymbol{\epsilon}_{t-\tau,\tau} + \cdots + \mathbf{B}^h\boldsymbol{\epsilon}_{t-h\tau,\tau} \tag{A.15}$$

$$\mathbf{e_{t+\tau}} = \boldsymbol{\mu_e} + \sum_{h=0}^{t/\tau} e^{-\boldsymbol{\Theta}h\tau}\boldsymbol{\epsilon}_{t-h\tau,\tau} \quad \Rightarrow \quad \mathbb{E}[\mathbf{e}_\infty] = \boldsymbol{\mu_e} \tag{A.16}$$

Regression residuals are 'mixed integrals' over the Brownian Motion (Ito Integrals).

$$\boldsymbol{\epsilon}_{t,\tau} \equiv \int_t^{t+\tau} e^{\boldsymbol{\Theta}(s-\tau)}\mathbf{S}d\mathbf{W_s} \tag{A.17}$$

The solution for $\mathbf{e_{t+\tau}}$ can be expressed as an integrated process over the residuals:

$$\mathbf{e_{t+\tau}} = \left(\mathbf{I} - e^{-\boldsymbol{\Theta}\tau}\right)\boldsymbol{\mu_e} + e^{-\boldsymbol{\Theta}\tau}\mathbf{e_t} + \int_t^{t+\tau} e^{\boldsymbol{\Theta}(s-\tau)}\mathbf{S}d\mathbf{W_s} \tag{A.18}$$

$$\mathbf{e_{t+\tau}} = \left(\mathbf{I} - e^{-\boldsymbol{\Theta}(t+\tau)}\right)\boldsymbol{\mu_e} + e^{-\boldsymbol{\Theta}(t+\tau)}\mathbf{e_0} + \int_0^{t+\tau} e^{\boldsymbol{\Theta}(s-(t+\tau))}\mathbf{S}d\mathbf{W_s} \tag{A.19}$$

If $\tau \to 0$ the process is dominated by residuals making it the Brownian Motion with $\boldsymbol{\epsilon}_{t,\tau} \sim N(\tau\boldsymbol{\Theta}\boldsymbol{\mu_e}, \tau\boldsymbol{\Sigma}_{OU})$ [4].

The solution for the covariance of VAR(1) residuals $\boldsymbol{\Sigma_\tau} = \mathbb{C}\mathrm{ov}[\boldsymbol{\epsilon_t, \epsilon_{t+\tau}}]$ relates to the OU scatter (dispersion coefficient) $\boldsymbol{\Sigma_{OU} = SS'}$ as

$$\boldsymbol{\Sigma_\tau} = (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} - e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})\tau}\right) \boldsymbol{\Sigma_{OU}} \qquad \Rightarrow \qquad \sigma_{OU} = \sqrt{\frac{2\,\theta}{1-e^{-2\theta\tau}}\,\mathbb{V}ar[\epsilon_{t,\tau}]} \tag{A.20}$$

$$\sigma_{eq} \approx \frac{\sigma_{OU}}{\sqrt{2\theta}} = \sigma_{OU}\sqrt{\frac{\tau}{2}} \tag{A.21}$$

The actual P&L from mean-reversion must be proportional to standard deviation σ_{eq} of equilibrium e_t. σ_{eq} is obtained from the model but a trade can be controlled by comparing the empirical dispersion of the P&L to σ_{eq}.

$$N(\boldsymbol{\mu_e}, (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1}\,\boldsymbol{\Sigma}_{OU}) \qquad \Rightarrow \qquad N(\mu_e, \sigma^2_{eq}) \tag{A.22}$$

We are interested in τ that is quite large compared to the period of data frequency. Therefore, the easiest assumption $\tau \to \infty$ provides an unconditional distribution (A.22) to make a probability statement about the reversion.

A.7 OU Process and Equilibrium Correction (VECM)

The following comparison shows that **the Ornstein-Uhlenbeck process fit** (A.11) **matches the features of equilibrium-correction term, which is commonly present in econometric cointegration models**.

$$\begin{aligned}
\mathbf{e_{t+\tau}} &= \left(\mathbf{I} - e^{-\boldsymbol{\Theta}\tau}\right)\boldsymbol{\mu_e} + e^{-\boldsymbol{\Theta}\tau}\mathbf{e_t} \\
\mathbf{e_{t+\tau}} - \mathbf{e_t} + \mathbf{e_t} - e^{-\boldsymbol{\Theta}\tau}\mathbf{e_t} &= \left(\mathbf{I} - e^{-\boldsymbol{\Theta}\tau}\right)\boldsymbol{\mu_e} \\
\Delta\mathbf{e}_{\mathbf{t},\tau} + (\mathbf{I} - e^{-\boldsymbol{\Theta}\tau})\mathbf{e_t} &= \left(\mathbf{I} - e^{-\boldsymbol{\Theta}\tau}\right)\boldsymbol{\mu_e}
\end{aligned}$$

$$\Delta\mathbf{e}_{\mathbf{t},\tau} = -\left(\mathbf{I} - e^{-\boldsymbol{\Theta}\tau}\right)\ \underline{(\mathbf{e_t} - \boldsymbol{\mu_e})} \tag{A.23}$$

compare wtih bivariate ECM A.4

$$= -(1-\alpha)\,(y_{1,t} - \beta y_{2,t} - \mu_e)$$

and multivariate case 2.8 with $\boldsymbol{e_t = \beta' Y_t}$

$$= \boldsymbol{\alpha}\ \underline{(\boldsymbol{\beta}'\mathbf{Y_t} - \boldsymbol{\mu_e})} \tag{A.24}$$

Introduction of the equilibrium level $\boldsymbol{\mu}_e$ means that we model $\Delta\mathbf{e}_{\mathbf{t},\tau}$. Otherwise that level should be moved outside of the cointegrating relationship and seen as a constant growth rate. Traditional VECM can be specified as $\Delta\mathbf{Y}_{\mathbf{t},\tau} = \boldsymbol{\Gamma}\Delta\mathbf{Y}_{\mathbf{t-1},\tau} + \Delta\mathbf{e}_{\mathbf{t},\tau} + \boldsymbol{\epsilon}_{t,\tau}$.

To show why the cointegrated residual (the spread that one would trade) is autoregressive of order one, AR(1) process, let's return to the cointegrated VAR(1) model with no intercept in VECM form

$$\begin{aligned}
\Delta \mathbf{Y_t} &= \boldsymbol{\alpha}\boldsymbol{\beta'}\boldsymbol{Y_{t-1}} + \boldsymbol{\epsilon}_t \qquad \text{(A.25)}\\
\mathbf{Y_t} &= \mathbf{Y_{t-1}} + \boldsymbol{\alpha}\boldsymbol{\beta'}\boldsymbol{Y_{t-1}} + \boldsymbol{\epsilon}_t\\
\underline{\boldsymbol{\beta'}\boldsymbol{Y_t}} &= (1 + \boldsymbol{\beta'}\boldsymbol{\alpha})\, \underline{\boldsymbol{\beta'}\boldsymbol{Y_{t-1}}} + \boldsymbol{\beta'}\boldsymbol{\epsilon_t}
\end{aligned}$$

which becomes

$$\underline{u_t} = \phi \underline{u_{t-1}} + \upsilon_t \qquad \text{(A.26)}$$

Adding extra drift terms, such as $\boldsymbol{\mu_{0,\tau}}$, will create a constant with no change in the nature of AR(1) process for the cointegrating residual, subject to the stability condition

$$|1 + \boldsymbol{\beta'}\boldsymbol{\alpha}| < 1 \quad \Rightarrow \quad \boldsymbol{\beta'}\boldsymbol{\alpha} < 0.$$

B Appendix B

Two quick examples were prepared to show the typical problems of the use of cointegration analysis to generate the mean-reverting spread from a pair of equities. Time period is over 100 trading days starting 23 October 2012 and ending 31 January 2013. The equities come from 'online industry' large-cap growth stocks: Amazon vs. Ebay pair has exposure to very similar *business risk factors*, while the Google vs. Apple pair is similar in its exposure to *market risk factors*, both are the top constituents of S&P 500.

This bivariate approach to cointegration is in the spirit of Engle-Granger procedure: select two I(1) time series with a common factor and test the spread of cointegrating relationship for stationarity. Augmented Dickey-Fuller test is a common test for unit-root, which is the inverse of stationarity. Any I(1) series can be represented as an *integrated series* (A.5). The illustrations show that even with very significant unit root tests on time series individually, the residual of the static equilibrium regression is not guaranteed to be unit root-free (Google vs. Apple). In turn, the stationarity of the spread (Amazon vs. Ebay) does not guarantee the quality of mean-reversion, i.e., trading opportunities. The low number of entry/exit points on boundary makes the arbitrageurs to exercise the half-life trading rule (i.e., enter/exit on the mean level).

The key problem from a viewpoint of a trader is that **the spread stays above or below its supposed long-term mean level over prolonged periods. The behaviour is common to the near-cointegrated situations**, in which the equilibrium is prone to shifting. The unit root-based tests suggest cointegration but the spread is not quite suitable for trading, which is evident without a further fitting of the spread to the Ornstein-Uhlenbeck process. This situation comes up frequently for equity pairs. Therefore, candidates for pairs trading cannot be selected on the sole basis of the econometric tests for cointegration.

A frequently asked question is for how long the estimates of cointegrating ratios $\boldsymbol{\beta}'_{\boldsymbol{Coint}}$ are likely to stay valid. Practitioners' advice cited in several sources is to estimate using one year of historic data and trade the estimates for a six-month period. An econometrican would construct a recursive estimate with confidence interval bounds [1].

B.1 Near-Cointegration: Google vs. Apple

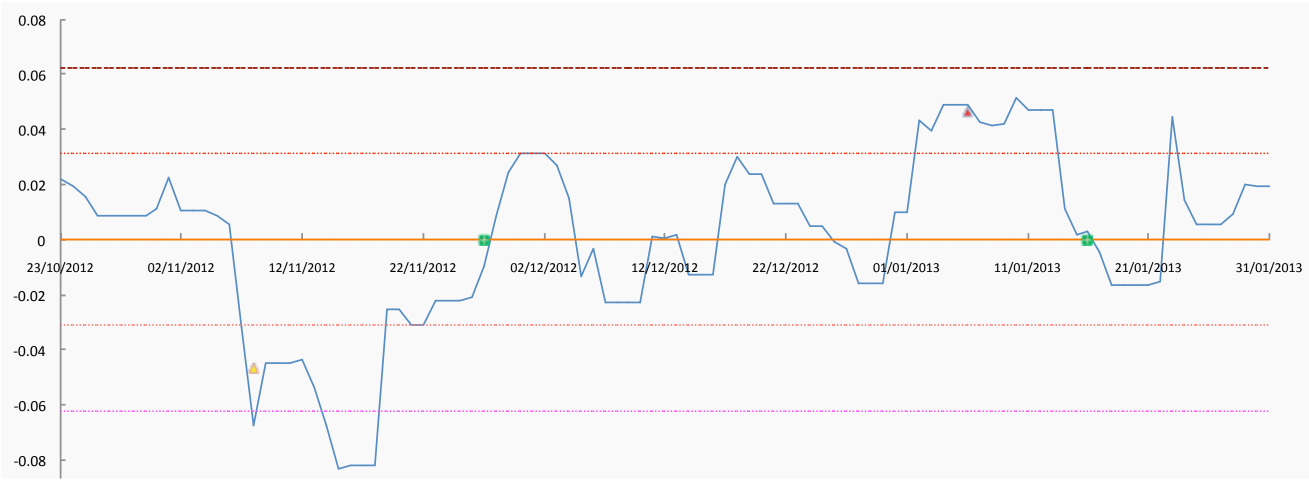

Figure B.1: Mean-reversion in a non-stationary spread e_t

	ADF t-stat	Unit Root
Google	-8.3564	Yes
Apple	-9.7552	Yes
Spread e_t	-2.7529	Not stationary

Table B.1: Cointegration Testing: Google vs. Apple (near-cointegration)

The static equilibrium regression on $\log Y_t$ has $R^2 = 0.4435$ and $\beta = -0.36$ (negative correlation). ADF test 5% Critical Value is -3.4559 (zero lagged values used).

B.2 Cointegration: Amazon vs. Ebay

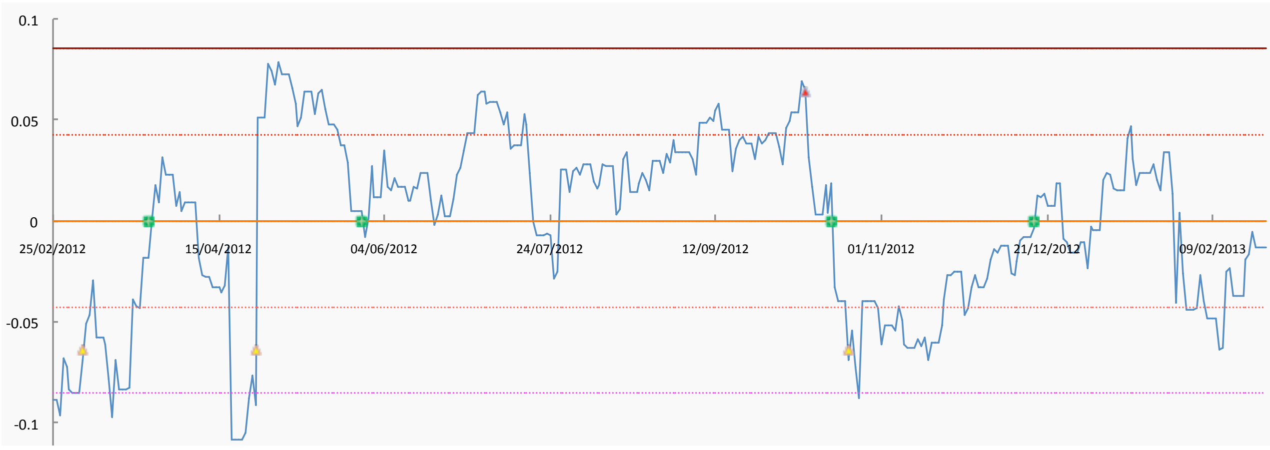

Figure B.2: Mean-reversion in a cointegrated spread e_t

	ADF t-stat	Unit Root
Amazon	-10.2395	Yes
Ebay	-10.4555	Yes
Spread e_t	-3.4979	Stationary

Table B.2: Cointegration Testing: Amazon vs. Ebay (formal cointegration)

The static equilibrium regression on $\log Y_t$ has $R^2 = 0.86$ and $\beta = 0.7668$ (strong positive correlation). ADF test 5% Critical Value is -3.4223 (three lagged values used).

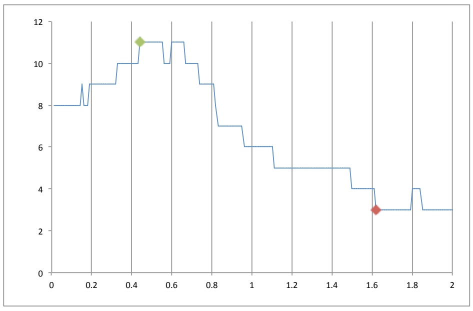

Figure B.3: Number of trades (vertical axis) as a function of boundary (in σ)

The tighter the boundaries above and below the mean exit point, the higher the number of trades and lower the holding period (less risk). However, trading tighter boundaries means less profit per trade. An optimisation can be performed in order to find the boundary that maximises the P&L.

Source: the charts were obtained using a template from the XLTP Bloomberg library.