Learning and Trusting Cointegration in Statistical Arbitrage

Richard V. Diamond
Fitch Learning, London, e-mail: r.diamond@cqf.com

Abstract
This technical paper offers simple views and uses of cointegration analysis which are not often clear in textbook presentations and research literature. Decompositions of time series and derivations, usually omitted in the presentation of equilibrium-correction models, are collected in a structured Appendix A. Instead of constructing a long-term forecast and entering what essentially is a trend-following strategy, the experienced arbitrageur would use the mean-reversion feature of the highly autoregressive spread generated by a cointegrated relationship. The suitability of mean-reversion for arbitrage is evaluated by fitting the spread to the Ornstein–Uhlenbeck process that provides a balanced fit for the features of correction to the long-run equilibrium. Conclusions formulate and address three common complaints about econometric tests for cointegration and instability of the equilibrium.

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

1 Linear simplicity of cointegration model
Cointegration is known as a model for a long-term relationship, which is a shortcut taken for 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 from shortcomings of 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 has 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 students 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 the unit root that makes a 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 to learning. In this paper, I guide the reader through what is called a "model specification" for cointegration analysis, while collecting the requisite transformations and hints in an accessible way. I aim to show how cointegration can be an intuitive and attractive concept for the design of spread trades.

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 analysis of the 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 the representation of series as a sum of shocks and utilizes 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 as to eliminate the common noise and produce a stationary spread ("a residual" of the cointegrated relation):

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

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 normalize the cointegrating relationship to $\beta'_{Coint} = [1, -\tilde{\beta}_2, \ldots, -\tilde{\beta}_n]$ – in this way, cointegration looks like differencing. The Granger–Johansen representation (15) 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*.[1]

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 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 the design simplicity of the cointegration model escapes first-time students. It is common that an analytical structure appears clear and simple to experts while being complex to novices who do not recognize 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 and x_t, the stochastic trend in each will not allow us to obtain a good model:

$$y_t = a + bx_t \tag{2}$$

We proceed to use the regression to model what we can, the growth rate Δy_t. How we move from a regression for y_t to this fully equivalent regression for Δy_t is explained by derivations (12) and (13) in Appendix A:

$$\Delta y_t = \beta_g \Delta x_t - (1-\alpha)\left[y_{t-1} - a_e - b_e x_{t-1}\right] \tag{3}$$

$$\Delta y_t = \beta_g \Delta x_t - (1-\alpha) e_{t-1} \tag{4}$$

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 recognize this as the steady-state model (we have a stationary spread series with a mean). The speed of correction toward the equilibrium level $-(1-\alpha)$ is an inferred (calibrated) parameter.

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

2.2 Laborious model specification

Issue 1: Model specification

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 a 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 to provide an explanation competes with the aim to condense as much information as possible. In the end, the latter wins: time series transformations become omitted, notation under-explained, and model explanations over-connected. In order to acquire a *threshold concept* of what cointegration is, we need actionable knowledge of (a) vector autoregression and (b) hypothesis testing – the 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 you need a quick computational introduction, [2] is the sole practical guide with code to estimate the cointegrating relationship matrix, in which maximum likelihood derivations are 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

The cointegrating relationships matrix $\Pi = \alpha\beta'$ is calculated *as if* it is factorized into cointegrating vectors β and speeds of adjustment to the equilibrium α. The factorization is not unique and requires normalization: $\beta_{MLE,C}$ opens up the space spanned by many cointegrating relationships; without "identifying restrictions," there could be no way of making sense of differencing weights/allocations/hedging ratios provided by cointegrating vectors. But somehow, we have to know the sensible restrictions on what are otherwise uncertain parameters.

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

$$\begin{aligned} \text{Estimate } \beta_{MLE} \;&\Rightarrow\; \text{Normalize } \beta_{MLE,C} \;\Rightarrow\; \text{Calibrate } \alpha \\ &\Rightarrow\; \text{Reconstruct } \Pi_{MLE} \;\Rightarrow\; \alpha\beta_{MLE,C}{}' \end{aligned}$$

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: 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 him the Nobel Prize for advancing the econometric analysis of non-stationary processes.

The key technical (matrix algebra) condition is for the relationships matrix $\Pi_{n\times n}$ to have a reduced rank of $r < n$ linearly independent rows. Otherwise, in model (17) (expressed in matrix form), the stationary $\Delta\mathbf{Y}_t$ would be equal to the non-stationary ΠY_{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.[2] Financial quants would recognize the non-uniqueness of factorization $\alpha\beta'$ that stems from the reduced rank as a case of uncertain parameters – exact values unknown but their range or space defined.

Issue 3: Deterministic trends

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

$$dY_{t,\tau} = \underline{Y_t \mu dt} + Y_t \sigma dW_t \Rightarrow \Delta \mathrm{Y}_t = \Gamma \Delta \mathrm{Y}_{t-1} + \underline{\alpha\beta' Y_{t-1}} + \underline{\mu_1 t} + \underline{\mu_{0,\tau}} + \epsilon_{t,\tau} \tag{5}$$

This blend of cointegration with other trends is an example of the most complex model specification of the Johansen framework for multivariate cointegration analysis (17). In addition to the equilibrium-correction term, it incorporates growth with time and constant growth trends [in (5), all three terms go in that order and are underlined].

Model-wise, the trends can be folded into the cointegrating relationship by looking for $\mu_1 t + \mu_0 = \alpha\rho_1 t + \alpha\rho_0$, 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 [1–3].

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

Monte-Carlo experiment

In order to check a number of assumptions about how trends can affect the specification of a cointegration model, I ran 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 a time-dependent trend were present but separated; in those models, coefficient μ_1 was extremely small, *less than other coefficients by* $O(10^{-2})$. Specifications with a time-dependent trend folded into a cointegrated relationship $\mu_1 t = \alpha\rho_1 t$ identified the least number of such relationships.[3] 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 discretized 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 $Y_t\mu dt$ to the shock $Y_t\sigma dW_t$. At any particular time step the shock term is likely to be much larger than the drift term. In the model for ΔY_t, the base GBM drift μdt corresponds to the constant growth rate $\mu_{0,\tau}$.[4] The drift is likely to interfere with the stationarity of the spread because it has to be folded in a cointegrating relationship for the VECM discretization to match the GBM model. This restriction comes from a small-time, low-volatility approximation of returns as follows:

$$\frac{Y_t - Y_{t-\tau}}{Y_{t-\tau}} \approx \mu\tau \;\Rightarrow\; \Delta Y_t = \mu\tau Y_{t-1} + \cdots \tag{6}$$

$$\Delta Y_t = \mu\tau Y_{t-1} + \alpha\beta' Y_{t-1} + \cdots = \alpha\rho_{0,\tau} Y_{t-1} + \alpha\beta' Y_{t-1} + \cdots \tag{7}$$

$$\mathbb{E}[\beta' Y_{t-1}] = \mu_e \;\Rightarrow\; \mathbb{E}[(\rho_{0,\tau} + \beta')Y_{t-1}] = \mu_e \tag{8}$$

Estimation $\rho_{0,\tau}$ is subject to the effectiveness of the earlier calibration of α. The restriction of the cointegration model (8) means adding to each of the cointegrating weights β' a quantity that is supposed to scale with time continuously (i.e., to be a constant for each length of time period). Non-constance of the quantity $\rho_{0,\tau}$ will make econometric tests fail in detecting cointegration among GBM series. Non-robust (changing) estimates for the cointegrating weights β' and equilibrium level μ_e pose a challenge for trade design.

We considered the underlined terms in (5), but there are effects coming from the Γ and $\varepsilon_{t,\tau}$ 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 Γ 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 discretization of a stochastic process is not a recipe for a flexible model and out-of-sample usefulness of that model.

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

$$\sum \epsilon_t \;\Rightarrow\; \sum dW_t \;\Rightarrow\; \epsilon_{t,\tau} \equiv \int_t^{t+\tau} Y_t \sigma dW_s \quad \sim N(0, \Sigma_\tau)$$

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

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

2.3 Limited forecasting

Issue 4: Design

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 to detect 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 $\mathbf{Y}_t$, we need to undo the differencing that is by definition an information-losing 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 $y_{1,t}$ depending on $y_{2,t}, \ldots, y_{n,t}$ comes at the price of making the forecast conditional on how well known the behavior of the independent variables is.

The model is limited to apply over the long run – the time period over which the correction toward the equilibrium is realized 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 ($\alpha \ll 0.1$) and addition of the extra terms ΔY_{t-k} or deterministic trends to the model for ΔY_t "erases" the impact of the equilibrium correction term $\beta' Y_{t-1}$. In a model with constant growth, the short-term forecast quickly converges to that constant $\mu_{0,\tau}$:

$$\Delta \mathrm{Y}_t = \mu_{0,\tau} + \alpha[\beta' Y_{t-1} - \mu_e] + \epsilon_{t,\tau} \;\Rightarrow\; \mathbb{E}[\Delta \mathrm{Y}_t] = \mu_{0,\tau} \tag{9}$$

If the equilibrium level shifts, $\mu_e \to \mu_e^*$, then any model that relies on equilibrium correction will provide *a forecast in the opposite direction*. Given that non-stationary time series are prone to shifts from 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, to 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, we can be in for a surprise as 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 modeling returns as

>

i.i.d. invariants. In efficient markets, it must not be feasible to extract information from returns.

To arrive at a forecasting equation for $\hat{Y}_t$, we have to go through the labor 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:

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

where $\hat{\Pi}$ is *not* limited to cointegrating relationships $\alpha\beta_{MLE,C}{}'$ but estimated by the generic vector autoregression. The powering up of the relationship matrix $\hat{\Pi}^h$ for each time step h multiplies the uncertainty of the forecast [1]. The further the projection, the wider its confidence interval bounds. Second, the outcome of long-term forecasting is trivial and known: growth at a constant rate, which can be seen as linear and time-dependent. But a trend-following strategy based on such a forecast is extremely simplistic and prone to shocks. Asymptotic analysis of distributions produced by non-stationary time series suggests that the 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 relations.

3 Constructing a spread trade

It is common for the use and usefulness of a model to depend on which of two sides are using it. We must realize 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 in 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.

3.1 Cointegration for hedgers

Hedgers and forecasters desire 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.[5] The focus shifts to the elaborate model specification and efforts depend on stationarity and non-heteroskedastic variance (i.e., they 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 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 more severe a correction is expected. That might work well for the low-frequency data, but it is unrealistic to expect a full 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 look like an unexpected accident. There are problems with the hedgers' world view.

3.2 Cointegration for traders

A trader should seek to utilize the qualities and phenomena that a model brings to light, such as the mean-reverting nature of the cointegrated spread $e_t = \beta' Y$. The spread is referred to as a "disequilibrium error" and gets thrown out in any forecasting. However, the arbitrageur's interest is not necessarily improvement of a forecast by fine-tuning the model specification but seeking to detect a particularly large "disequilibrium error" while minimizing the model error. The only "guarantee" a trader needs from the cointegration analysis is that it will deliver the mean-reversion.

A cointegration model can be seen as an affine filter applied to identify a mean-reverting stationary spread (if it exists). This application satisfies the *Financial Modelers' Manifesto* [9] in several ways. Going for a theory-free exploitation of a phenomenon in time series is a bold move, but the elegance of the equilibrium-correction idea does not contradict the reality of different asset classes and markets. The empirical attributions that can emerge from the uses and alterations of the model are more useful to traders than an elaborate pre-specified model. The *Manifesto* can be extended with one more claim: practitioners should demand affine solutions, for which it is possible to obtain explicit no-arbitrage conditions. *The cointegrated relationship defined by weights β' that deliver $\mathbb{E}[\beta' Y_t - \mu_e] = 0$ represents such a no-arbitrage condition.*

To design and manage a trade (i.e., control entry and exit, estimate P&L and drawdowns), we require the following items of information:

1. *Weights* for a set of instruments to construct a spread trade.
2. *Speed of mean-reversion* to evaluate feasibility of the trade.
3. *Half-life: time between the perfect equilibrium situations* is the third important piece of information that is not provided by econometric analysis.

The weights do not need to come from a formal cointegration model. It is possible to check whether there is any significant speed of adjustment toward the equilibrium α for any set of weights. This problem of non-unique factorization that haunts cointegration parameters is the problem of hedgers, who would like to attribute economic meaning to the allocations and equilibrium level. Traders should be interested in *the quality of mean-reversion*, a broader 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 the mean level μ_e.

The spread $e_t = \beta' Y_t$ is a special AR(1) process without a unit root $I(0)$ (32). These properties make it possible *to evaluate the quality of mean-reversion by fitting the spread to the Ornstein–Uhlenbeck process and examining the parameters:* mean gives the equilibrium level μ_e, standard deviation σ gives the upper and lower bounds, and the speed of mean-reversion ("recall force") θ indicates its quality without the need to conduct formal econometric cointegration tests on data Y_t. Theta can also be converted to the half-life, as $\tilde{\tau} \propto \frac{\ln 2}{\theta}$ [4]. The fitting recipe is surprisingly simple and uses a VAR(1) model (21).

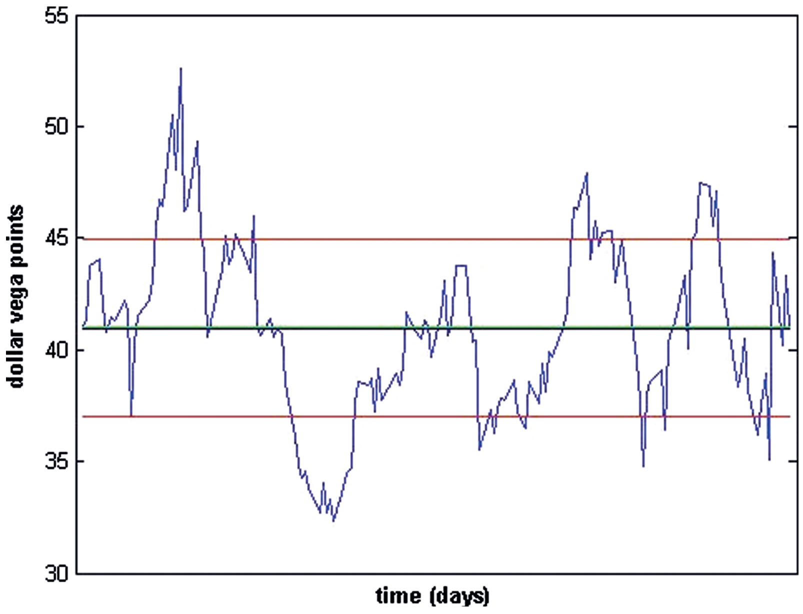
In comparison, there is a technical contribution of the Ornstein–Uhlenbeck SDE solution to the econometric parameterizations of the spread as an equilibrium-correction term (30). It is insightful to find how the parameters fall into their correct places with identifiable roles: in the bivariate case, *the speed of adjustment toward the equilibrium is proportional to the autocorrelation of the mean-reverting process,* $\alpha \propto e^{-\theta\tau}$.

Mean-reversion begins with $\theta > 0$. The higher the theta, the shorter the half-life of reversion to the mean, and the better the quality of mean-reversion becomes. For the high-frequency data of the instruments linked by a term structure (i.e., swaps and futures), it is not uncommon to observe θ=10, … ,200 and above.

3.3 An example

An example of the stationary and mean-reverting spread $e_t \sim I(0)$ was obtained from a cointegrating relationship among four futures linked to a volatility index.

Figure 1: Mean-reversion in a cointegrated spread e_t, $\theta = 42.39$.



The observation sample covered 142 trading days from March to September 2012. Figure 1 shows the spread fitted to the Ornstein–Uhlenbeck process with mean and 1σ bounds.

We see that the cointegration model acts as a filter: *for daily data, the spread has a longer half-life.* In this case, the mean-reversion occurs ca. every 22 days. Trading half-life: entering on a bound and exiting on the mean would be the default arbitrage strategy, but if we would like to catch a bigger move, the trade life will be longer. Execution of trades with long life in terms of "market time" is a bold move that requires conviction and position management. However, there is an advantage in having a horizon that is above the near-term horizons of other market participants (i.e., one-month implied volatility).

3.4 The near-cointegrated situation

The near-cointegrated situation 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 w.r.t. the mean. This indicates that the equilibrium level is prone to shifting. Approaching the situation, there are two further distinctions: (2) the spread mean-reverts but trends (up or down), and (3) the spread "spends" some time around the mean. In both cases, it is still possible to observe the mean-reversion w.r.t. 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. This 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 us to construct trades with conviction even if cointegration is fractionated and the spread is ill-behaved.

4 Trusting cointegration

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

The cointegration model reveals a deterministic trend of a specific kind: equilibrium correction over the long run. Time series decompositions (11) and (15), as well as models (13) and (20), demonstrate that it is possible to remove the stochastic trend from an integrated time series and construct a stationary as well as an 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) $\beta'_{Coint}\,Y$ 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 in 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"? We 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 those 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.

4.1 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. At the other end of the 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 time series simulated as a pair (i.e., returns of one asset leading to 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 the 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., normalized to start at the same point). It follows that testing all combinations of marketable assets for cointegration is neither a computationally efficient nor reliable approach.

4.2 Addressing complaint 2

Sometimes the complaint 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. We just have 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 e_t is presented in Appendix B. The fitting is surprisingly simple, by estimation of VAR(1) only. The match between the regression and the O–U process parameters is affine, see (22) and (23). For cointegration to exist, the technical condition is that *some* eigenvalues of transition matrix Θ have positive real parts [4]. If that fairly relaxed numerical condition is satisfied, the spread $\mathbf{e}_t$ does converge to its deterministic drift μ_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-rate instruments. In this case, the modeling has to go back to first principles.

The cointegration model offers a decomposition into one stochastic trend (in general, the Levy process) *and several deterministic trends*. The avant-garde of econometric research took turns in a technical attempt to address the shortcomings of cointegration by *modeling of the stochastic trend with fractional Brownian motion*, which is a continuous-time model for long memory [8].

The autocorrelation of fractional Brownian motion decays according to the power law τ^{2d-1}, which is slower than the exponential pattern of Ornstein–Uhlenbeck $e^{-\theta\tau}$. The dual nature of fractional Brownian motion allows us to model integrated series $I(1)$ by setting parameter $d \approx 0$ as well as more stationary-like series with low values of the Hurst exponent $H < 0.2$.[6] Interest-rate swaps with pronounced long memory are modeled with $H > 0.5$ as the change increments not fully independent invariants (think of a bond getting closer to maturity). The use of fractionally integrated processes promises to address the problem of the shifting equilibrium. Simulated fractionally cointegrated time series exhibit different degrees (strengths) of cointegration at different time periods.

4.3 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 stabilize the equilibrium) by 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 the 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 gray cases which challenge the econometric tests that ultimately rely on closed-form decompositions and the equilibrium-correction term (13). The term is usually small but significant, which confirms at least the conditional equilibrium.[7] Its impact is mediated by the calibrated speed of adjustment parameter α.

The small impact of equilibrium correction, which adds up over the long run, can be easily disturbed by short-term shocks. While the theory suggests that it would take a while to correct such shocks, *the market correction is commonly observed as a 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 were a constant scalable with time as assumed by geometric Brownian motion, there would be no technical problem of 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 such 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 a multivariate cointegrated system, standardized w.r.t. its transition speed Θ, shows how the shift toward 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 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 on the generic no-arbitrage interpretation (w.r.t. 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 is 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-i.i.d. movements[8] and delivers a robust spread that is still of high frequency. The model provides that the spread is more likely to remain stationary than the underlying time series.

Cointegration must be utilized in the spirit of data exploration: with attention to detail in model specification but without preoccupation with statistical testing *per se*. Fischer Black noted that significance tests themselves are almost of no value in "the real world of research," and the best approach would be "to 'explore' a model" [10]. This I interpret as exploring the behavior of model parameters under varying assumptions about deterministic trends and pre-processing of data (i.e., standardizing, projecting in orthogonal dimensions, and sorting cointegration candidates by fundamental criteria).

It is better to use an affine model in order to make data alive rather than blindly rely on complex specifications of statistical tests coded into software packages. The use of such pre-specified tests requires the preparation of data, which already assumes a lot about its internal structure and trends. Treating cointegration analysis as a set of statistical tests run on prepared data can turn the modeling effort into a medieval anatomy experiment with a low chance of finding something that has not already been specified.

Appendix A

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

A.1 Decomposing AR as an MA process

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

$$\frac{1}{1-\beta L} \approx 1 + \beta L + \beta^2 L^2 + \cdots \equiv \sum_{i=0}^{\infty} (\beta L)^i \qquad \text{for lag operator } L \tag{10}$$

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

$$y_t = \frac{\beta_0}{1-\beta_1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_1^2 \epsilon_{t-2} + \cdots \qquad \text{s.t. } |\beta_1| < 1 \text{ gives stationary } y_t \tag{11}$$

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 the 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:

$$y_t = \alpha y_{t-1} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \epsilon_t \tag{12}$$

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

$$\begin{aligned}
y_t - \underline{y_{t-1}} &= \alpha y_{t-1} - \underline{y_{t-1}} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} - \underline{\beta_1 x_{t-1}} + \underline{\beta_1 x_{t-1}} + \epsilon_t \\
\Delta y_t &= -(1-\alpha) y_{t-1} + \beta_0 + \beta_1 \Delta x_t + (\beta_1 + \beta_2) x_{t-1} + \epsilon_t \\
&= \beta_1 \Delta x_t - (1-\alpha)\left[y_{t-1} - \frac{\beta_0}{1-\alpha} - \frac{\beta_1+\beta_2}{1-\alpha} x_{t-1} \right] + \epsilon_t \\
&= \beta_1 \Delta x_t - (1-\alpha)\left[y_{t-1} - a_e - b_e x_{t-1} \right] + \epsilon_t \\
&= \beta_1 \Delta x_t - (1-\alpha) e_{t-1} + \epsilon_t
\end{aligned} \tag{13}$$

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.

Models (12) and (13) are fully equivalent.

A.3 Granger–Johansen representation (perfect unit root)

We can represent a perfect unit-root process $x_t = x_{t-1} + \varepsilon_{x,t}$ as a stochastic process that allows different "trends" at any point in time:

$$x_t = \underbrace{\sum_{s=1}^{t} \epsilon_{x,s}}_{} + x_0 \tag{14}$$

Regressing another perfect unit-root process x_t on y_{t-1} and substituting:

$$y_t = x_{t-1} + \epsilon_{y,t} = \underbrace{\sum_{s=1}^{t} \epsilon_{x,s}}_{} - \epsilon_{x,t} + \epsilon_{y,t} + X_0 \tag{15}$$

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 (14) from (15) eliminates the common stochastic process and leaves the stationary autoregressive residual $e_t = \varepsilon_{y,t} - \varepsilon_{x,t} \equiv (y_t - b_e x_t - a_e)$.

In the multivariate setting, the joint stationary residual is written as $\mathbf{e}_t = \beta'_{Coint} \mathbf{Y}_t$ 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 $\alpha\mathbf{e}_{t-1}$ that corresponds to equilibrium correction in the bivariate case (13). Notice the change of convention $-(1-\alpha) \to \alpha$:

$$\Delta \mathrm{Y}_t = \Gamma \Delta \mathrm{Y}_{t-1} + \underline{\alpha \mathrm{e}_{t-1}} + \epsilon_t \tag{16}$$

In the econometrics literature, full VECM is often expressed as

$$\Delta \mathrm{Y}_t = \Gamma \Delta \mathrm{Y}_{t-1} + \underline{\Pi \mathrm{Y}_{t-1}} + \Psi \mathrm{D}_t + \epsilon_t \tag{17}$$

where

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

A.5 Reduced rank matrix

Granger representation theorem states that if matrix Π has a reduced rank $r < n$ then it can be factorized as

$$\Pi = \alpha\beta'$$

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

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

If we restrict $\Pi \propto \beta$ to $r = 2$ cointegrating vectors, the result is in line with (13):

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

The vector error-correction term for a whole system is

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

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

It makes sense to vectorize multiplication of coefficients α,β' when the relationship matrix Π is presented alone. For a 3×3 example with rank $r = 2$ and normalized weights,

$$\Pi = \mathrm{Vec}\left(\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \right) \mathrm{Vec}\left(\begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} \\ 1 & -\beta_{22} & -\beta_{23} \end{bmatrix} \right) \tag{18}$$

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

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

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

$$de_t = -\Theta(e_t - \mu_e)\,dt + S\,dW_t \tag{19}$$

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

$$e_{t+\tau} = \left(I - e^{-\Theta\tau}\right)\mu_e + e^{-\Theta\tau}e_t + \epsilon_{t,\tau} \tag{20}$$

The vector autoregression equivalent for a small time period τ is

$$e_{t+\tau} = C + Be_t + \epsilon_{t,\tau} \tag{21}$$

Once VAR(1) is estimated, we can solve for a square matrix Θ and vector μ_e:

$$e^{-\Theta\tau} = B \;\Rightarrow\; \Theta = -\frac{\ln B}{\tau} \tag{22}$$

$$\left(I - e^{-\Theta\tau}\right)\mu_e = C \;\Rightarrow\; \mu_e = \frac{C}{I - B} \tag{23}$$

If we use (11) to decompose the VAR(1) process (21) into its MA(∞) representation:

$$e_{t+\tau} = \frac{C}{I - B} + \epsilon_{t,\tau} + B\epsilon_{t-\tau,\tau} + \cdots + B^h\epsilon_{t-h\tau,\tau} \tag{24}$$

$$e_{t+\tau} = \mu_e + \sum_{h=0}^{t/\tau} e^{-\Theta h\tau}\epsilon_{t-h\tau,\tau} \;\Rightarrow\; \mathbb{E}[e_\infty] = \mu_e \tag{25}$$

It is interesting to trace what happens with the residuals of regression: they are "mixed integrals" over the Brownian motion (Ito integrals):

$$\epsilon_{t,\tau} \equiv \int_t^{t+\tau} e^{\Theta(s-\tau)}S\,dW_s \tag{26}$$

The scatter generator (dispersion coefficients) matrix S is the result of a decomposition $\Sigma = SS'$ obtained from the residual autocovariance matrix $\Sigma_\tau = \mathrm{Cov}[\varepsilon_t, \varepsilon_{t+\tau}]$ in [4].

The solution to the Ornstein–Uhlenbeck SDE has the scatter parameterized as

$$\sigma = s\sqrt{\frac{1 - e^{-\theta\tau}}{2\theta}} \quad\Rightarrow\quad \Sigma = (\Theta \oplus \Theta)^{-1}\left(I - e^{-(\Theta\oplus\Theta)\tau}\right)\Sigma_\tau$$

The unconditional autocorrelation plays the key part and has an exponential decay

$$\mathrm{Corr}[e_t, e_{t+\tau}] = e^{-\Theta\tau}. \tag{27}$$

The solution can also be expressed as an integrated process over the residuals:

$$e_{t+\tau} = \left(I - e^{-\Theta\tau}\right)\mu_e + e^{-\Theta\tau}e_t + \int_t^{t+\tau} e^{\Theta(s-\tau)}S\,dW_s \tag{28}$$

$$e_{t+\tau} = \left(I - e^{-\Theta(t+\tau)}\right)\mu_e + e^{-\Theta(t+\tau)}e_0 + \int_0^{t+\tau} e^{\Theta(s-(t+\tau))}S\,dW_s \tag{29}$$

As $t \to \infty$ the impact of residuals wanes, and the process converges to its long-term mean μ_e, while if $\tau \to 0$ the process is dominated by residuals, making it the Brownian motion with $\varepsilon_{t,\tau} \sim N(\tau\Theta\mu_e, \tau\Sigma)$ [4].

A.6.1 Contribution

Returning to the Ornstein–Uhlenbeck solution (20) without residuals:

$$\begin{aligned}
e_{t+\tau} &= \left(I - e^{-\Theta\tau}\right)\mu_e + e^{-\Theta\tau}e_t \\
e_{t+\tau} - e_t + e_t - e^{-\Theta\tau}e_t &= \left(I - e^{-\Theta\tau}\right)\mu_e \\
\Delta e_{t,\tau} + (I - e^{-\Theta\tau})e_t &= \left(I - e^{-\Theta\tau}\right)\mu_e \\
\Delta e_{t,\tau} &= -\left(I - e^{-\Theta\tau}\right)(e_t - \mu_e) \qquad (30)
\end{aligned}$$

$$\begin{aligned}
&= -(1-\alpha)(y_{1,t} - \beta y_{2,t} - \mu_e) \\
&= \alpha\,(\beta' Y_t - \mu_e) \qquad (31)
\end{aligned}$$

The comparison shows that *the Ornstein–Uhlenbeck process provides a balanced, if not perfect, fit for the features of the equilibrium-correction term* in cointegration models like the VECM $\Delta\mathbf{Y}_{t,\tau} = \Gamma\Delta\mathbf{Y}_{t-1,\tau} + \Delta\mathbf{e}_{t,\tau} + \varepsilon_{t,\tau}$. Introduction of the level μ_e means that we essentially model $\Delta\mathbf{e}_{t,\tau}$. Otherwise, the equilibrium level should be moved outside the cointegrating relationship and seen as a constant growth rate.

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

$$\begin{aligned}
\Delta Y_t &= \alpha\beta' Y_{t-1} + \epsilon_t \qquad (32)\\
Y_t &= Y_{t-1} + \alpha\beta' Y_{t-1} + \epsilon_t \\
\underline{\beta' Y_t} &= (1 + \beta'\alpha)\,\underline{\beta' Y_{t-1}} + \beta'\epsilon_t
\end{aligned}$$

which becomes

$$\underline{u_t} = \phi\underline{u_{t-1}} + \upsilon_t \tag{33}$$

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

$$|1 + \beta'\alpha| < 1 \;\Rightarrow\; \beta'\alpha < 0.$$

Appendix B

Two quick examples were prepared to show the typical problems of using cointegration analysis to generate the mean-reverting spread from a pair of equities. The time period is over 100 trading days starting October 23, 2012 and ending January 31, 2013. The equities come from "online industry" large-cap growth stocks: the 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 the S&P 500.

This bivariate approach to cointegration is in the spirit of the Engle–Granger procedure: select two $I(1)$ time series with a common factor and test the spread of the cointegrating relationship for stationarity. The augmented Dickey–Fuller test is a common test for the unit root, which is the inverse of stationarity. Any $I(1)$ series can be represented as an *integrated series* (14). 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 the boundary makes arbitrageurs exercise the half-life trading rule (i.e., enter/exit on the mean level).

The key problem from the viewpoint of a trader is that *the spread stays above or below its supposed long-term mean level over prolonged periods. This behavior is common to 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 econometric tests for cointegration.

A frequently asked question is for how long the estimates of cointegrating ratios β'_{Coint} are likely to stay valid. Practitioners' advice cited in several sources is to estimate using 1 year of historic data and trade the estimates for a 6-month period. An econometrican would construct a recursive estimate with confidence interval bounds [1].

B.1 Near-cointegration: Google vs. Apple

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

	ADF *t*-statistic	Unit root
Google	−8.3564	Yes
Apple	−9.7552	Yes
Spread e_t	−2.7529	Not stationary

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

B.2 Cointegration: Amazon vs. ebay

Table 2: Cointegration testing: Amazon vs. ebay (formal cointegration)

	ADF *t*-statistic	Unit root
Amazon	−10.2395	Yes
Ebay	−10.4555	Yes
Spread e_t	−3.4979	Stationary

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

Figure 2: Mean-reversion in a non-stationary spread e_t.

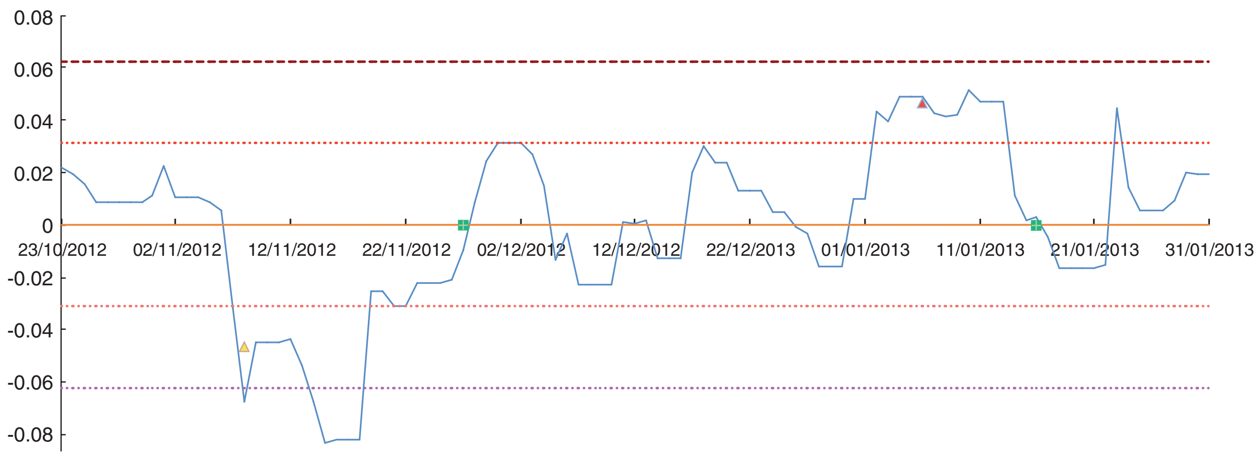


Figure 3: Mean-reversion in a cointegrated spread e_t.

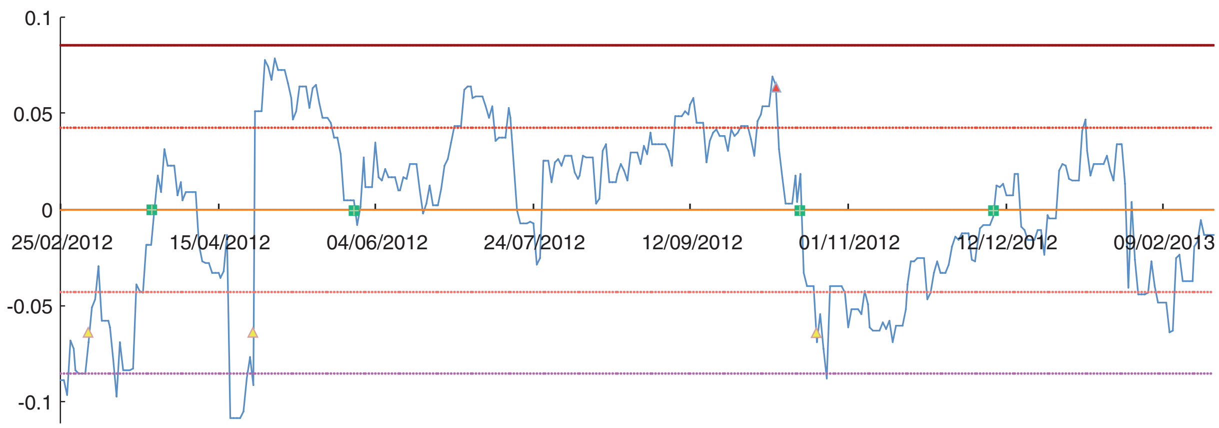
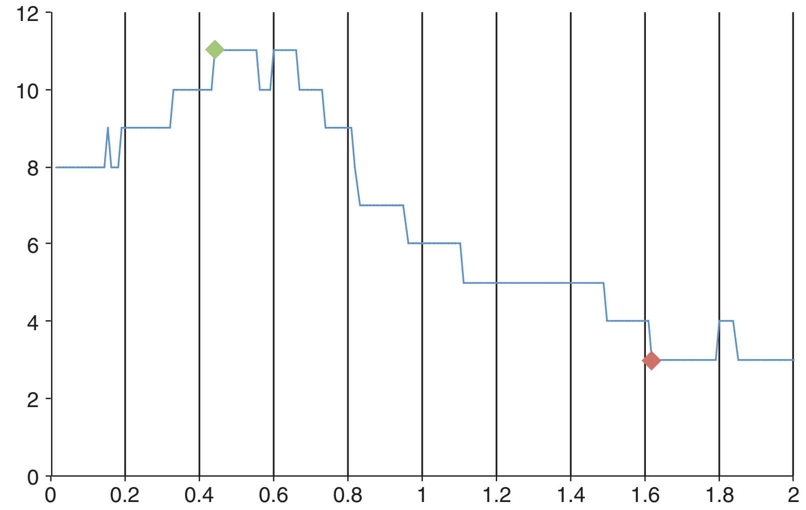


Figure 4: Number of trades (vertical axis) as a function of boundary (in σ).



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

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

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.

ENDNOTES

1. MA(∞) decomposition (11) is a theoretical exercise but it reveals the stochastic trend in a non-stationary time series. It also completes the quest for invariance: asymptotically, any noise is an independently and identically distributed process (in its increments), *an i.i.d. invariant* of a model. Once such an i.i.d. invariant is identified, it can be substituted with random numbers to generate projections, stress-tested, and, in the case of cointegration analysis, removed by the special differencing $\beta'_{Coint} Y_t$.
2. 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 we can take the opportunity to revisit matrix algebra and represent matrix calculations in an explicit form as done in (18).
3. Model *H** of MATLAB's *jcitest* function that implements the Johansen framework. The model choice is to restrict trends to a cointegrating relationship or add, e.g. as the extra terms in model (5) for $\Delta\mathbf{Y_t}$.
4. The expected growth $\mathbb{E}[Y_t|Y_0] = Y_0 e^{\mu_0, \tau^t}$ should not be confused with the $\mu_1 t$ term.
5. The next steps are heterogeneous regression with time-varying parameters and using space-state representation techniques, such as Kalman filtering. The "time-varying" term is misleading because the parameters are really state-varying. Because it is the equilibrium level μ_e that shifts, it makes sense to talk about equilibrium-varying or conditional equilibrium cointegration.
6. 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.
7. Otherwise, the term would be insignificant. The specification of the equilibrium-correction term is unequivocal in that it represents the long-run equilibrium.
8. Larger jumps are a consequence of order toxicity, which makes them pre-visible. They can also be modeled as structural breaks.

REFERENCES

[1] Hendry, D. and Juselius, K. Explaining cointegration analysis: Part I. 2001. *Energy Journal*, **21**: 1–42; Explaining cointegration analysis: Part II. *Energy Journal*, **22**: 75–120. See also *Teaching Undergraduate Econometrics at Oxford* at www.youtube.com/watch?v=We12iDwwYEE

[2] Jang, K. and Ogaki, M. 2001. *User Guide for Johansen's Method*. Available at www.scribd.com/doc/37684294/User-Guide-for-Johansen%EF%BF%BDs-Method.

[3] Zivot, E. and Wang, J. *Modeling Financial Time Series with S-PLUS*, 2nd edn. Berlin: Springer-Verlag, 2006. See specifically Chapter 12: VAR Models and Cointegration.

[4] Meucci, A. 2010. *Review of Statistical Arbitrage, Cointegration, and Multivariate Ornstein–Uhlenbeck*. 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] Damghani, M., *et al.* 2012. The misleading value of measured correlation. *Wilmott magazine*, Nov, 64–73.

[6] Harlacher, M. Cointegration based statistical arbitrage. Master's Thesis, Swiss Federal Institute of Technology, Zurich.

[7] Hendry, D. Professor Sir Clive W.J. 2010. Granger and cointegration. *Journal of Financial Econometrics*, **8**(2): 162–168.

[8] Johansen, S. and Nielsen, M. 2012. Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, **80**: 2667–2732.

[9] Wilmott, P. and Derman, E. *The Financial Modelers' Manifesto*. 2009. In P. Wilmott, Ed. *Frequently Asked Questions in Quantitative Finance.* New York: Wiley.

[10] Derman, E. 2009. On Fischer Black: Intuition is a merging of the understander with the understood. Talk given at Bloomberg, NY, Nov. 24.