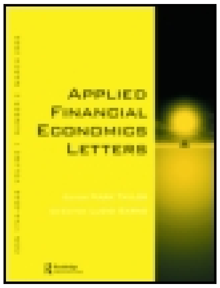


Applied Financial Economics Letters
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rafl20

# Demonstrating error-correction modelling for intraday statistical arbitrage

Brian Jacobsen [a]
[a] Business Economics, Wisconsin Lutheran College, Milwaukee, USA

Version of record first published: 08 Jul 2008

*Applied Financial Economics Letters*, 2008, **4**, 287–292

Routledge
Taylor & Francis Group

# Demonstrating error-correction modelling for intraday statistical arbitrage

Brian Jacobsen

*Business Economics, Wisconsin Lutheran College, Milwaukee, USA*
*E-mail: brian_jacobsen@wlc.edu*



Applying cointegration analysis to security price movements illustrates how securities move together in the long-term. This can be augmented with an error-correction model to show how the long-run relationship is approached when the security prices are out of line with their cointegrated relationship. Cointegration and error-correction modelling promises to be useful in statistical arbitrage applications: not only does it show what relative prices of securities should be, but it also illuminates the short-run dynamics of how equilibrium should be restored along with how long it will take. Cointegration, coupled with error-correction modelling, promises to be a profitable way of implementing statistical arbitrage strategies.[1] Bondarenko (2003) and Hogan *et al.* (2004) defined statistical arbitrage as an attempt to exploit the long-horizon trading opportunities revealed by cointegration relationships. Alexander and Dimitriu (2005) showed how cointegration is a better way of implementing a statistical arbitrage strategy than other conventional ways, like the use of tracking error variance minimization. These previous studies, however, did not add error-correction modelling to the trading strategies. This article seeks to fill that gap, by presenting how to implement a statistical arbitrage strategy based on cointegration and error-correction modelling.

## I. Theoretical Background

For time-series data, the concept of a 'mean' or a 'variance' becomes meaningless if the time-series is nonstationary. A time series is nonstationary if it is generated by a joint probability distribution that is not independent of time. For example, if a time series is given by $x_i, x_{i+1}, \ldots, x_{i+n}$ then this series is nonstationary if the joint probability distribution generating the series is not independent of $i$. In other words, the data-generating process changes over time.

A special type of nonstationary process, but by no means the only type of nonstationary process, is a process that has what is called a 'unit-root'. If the time series has a unit-root, that variable's use in regression analysis can give spurious results, showing a relationship among the levels of the variables without any real relationship existing. Additionally, the parameter estimates are inconsistent. Only when

[1] For example, see Kumar and Seppi (1994), Wang and Yau (1994), Forbes *et al.* (1999), Canjels *et al.* (2004), Tatom (2002), Harasty and Roulet (2000) and Laopodis and Sawhney (2002). They have applications ranging from index arbitrage to gold-point arbitrage during the pre-World War I era.

*Applied Financial Economics Letters* ISSN 1744–6546 print/ISSN 1744–6554 online  
http://www.tandf.co.uk/journals
DOI: 10.1080/17446540701720550

nonstationary variables are cointegrated can they be used meaningfully in regression analysis.

A variable that has a unit-root is also said to be 'integrated of order one' [denoted by I(1)], if first differencing the variable gets rid of the unit-root. If individual time series are I(1), they may be cointegrated where there exists one or more linear combinations of the variables that is stationary, even though the individual variables are not stationary. Two variables that are cointegrated never move 'too far away' from each other. A lack of a cointegrating relationship suggests there is no long-run link between the variables.

A series, $p_t$, has a unit-root if it follows a process like (1) with $\pi = 1$ ($\delta$ is a parameter and $e_t$ is an error term):

$$p_t = \delta + \pi p_{t-1} + e_t \tag{1}$$

This is what is typically known as a random walk and can be rewritten as

$$\Delta p_t = \delta + (\pi - 1)p_{t-1} + e_t \tag{2}$$

where $\Delta p_t = p_t - p_{t-1}$.

Dickey and Fuller (1979) have shown that if Equation 2 follows a unit root, then $\pi = 1$. If $\delta \neq 0$, then it has a unit root with a drift. The traditional $t$-test can be used to test the null hypothesis that the series has a unit root (Dickey *et al.*, 1991). A series is said to be integrated of order one if the first difference, $\Delta p_t$, is stationary, but the level is not stationary.

Cointegration is a long-term relationship between two time series that are both I(1), but a linear combination of the two series is I(0). If you have two time series, $p_t^1, p_t^2$, a linear relationship that is stationary can be represented as follows:

$$p_t^1 - \chi p_t^2 = \varepsilon_t \tag{3}$$

To be stationary, and for $\chi$ to be the cointegrating factor, $\varepsilon_t$ must have a constant mean, variance and autocovariance.

The simplest error-correction model, which captures the short-run dynamics of how the long-run equilibrium of Equation 3 is restored, is given as follows:

$$\begin{bmatrix}\Delta p_t^1\\ \Delta p_t^2\end{bmatrix} = \begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} + \begin{bmatrix}\beta_{1,1} & \beta_{1,2}\\ \beta_{2,1} & \beta_{2,2}\end{bmatrix}\begin{bmatrix}\Delta p_{t-1}^1\\ \Delta p_{t-1}^2\end{bmatrix} + \begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}\left(p_{t-1}^1 - \chi p_{t-1}^2\right) + \begin{bmatrix}\upsilon_t^1\\ \upsilon_t^2\end{bmatrix} \tag{4}$$

$\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}$, $\begin{bmatrix}\beta_{1,1} & \beta_{1,2}\\ \beta_{2,1} & \beta_{2,2}\end{bmatrix}$ and $\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}$ are parameters to be estimated. $\begin{bmatrix}\upsilon_t^1\\ \upsilon_t^2\end{bmatrix}$ is a vector of white noise. The vector $\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}$ represents how the preceding disequilibrium is corrected. One divided by each coefficient can be interpreted as how many time periods it takes for the disequilibrium to be corrected. The smaller this number (the larger, in magnitude, the coefficient), the more efficient the market for securities is, which means disequilibria are quickly arbitraged away.

Those securities with a cointegrating relationship have a long-run equilibrium relationship between their prices. Any deviation from this equilibrium should be corrected over time, which means an investment strategy can be based on identifying the disequilibrium: shorting the overpriced security and going long the under-priced security. An error-correction model can even identify how long this mispricing is likely to persist.

The error-correction term should be negative if there is a tendency for the prices to converge to a long-run relationship. If the error-correction term is positive, then the difference between the securities grows, and this is similar to an explosive relationship where the securities repel each other.

## II. Empirical Test

To illustrate this trading strategy I chose two exchange traded funds: iShares® S&P 500 Index Fund (ticker symbol IVV) and iShares® Russell 1000 Index Fund. Using exchange traded funds have the advantage that short selling is not restricted to being done only on an up-tick. Additionally, there is a significant overlap in the coverage of these two funds, so even without a statistical test to prove it, common sense suggests these funds should tend to move together. Because of the heavy trading in these funds, the market is very deep and liquid. For purposes of this article, I have used minute-by-minute data from 1 March 2006 to 19 December 2006. Reuters® provided all the data.

The first step is to show that each funds price series has a unit root. Figures 1 and 2 are the price series for these two funds and they both have a unit root. The second step is to show that they are cointegrated. For this it is a mere test of whether the residuals from regressing one fund on the other have a unit root. Figure 3 shows the residual series and this series does not have a unit root; hence, IWB and IVV are cointegrated.

There were a total of 20 591 observations on these two funds. The cointegrating relationship between these two funds is given by IVV = –1.9148 + 1.87224 IWB. The coefficient of determination between the price series is 99.704%.

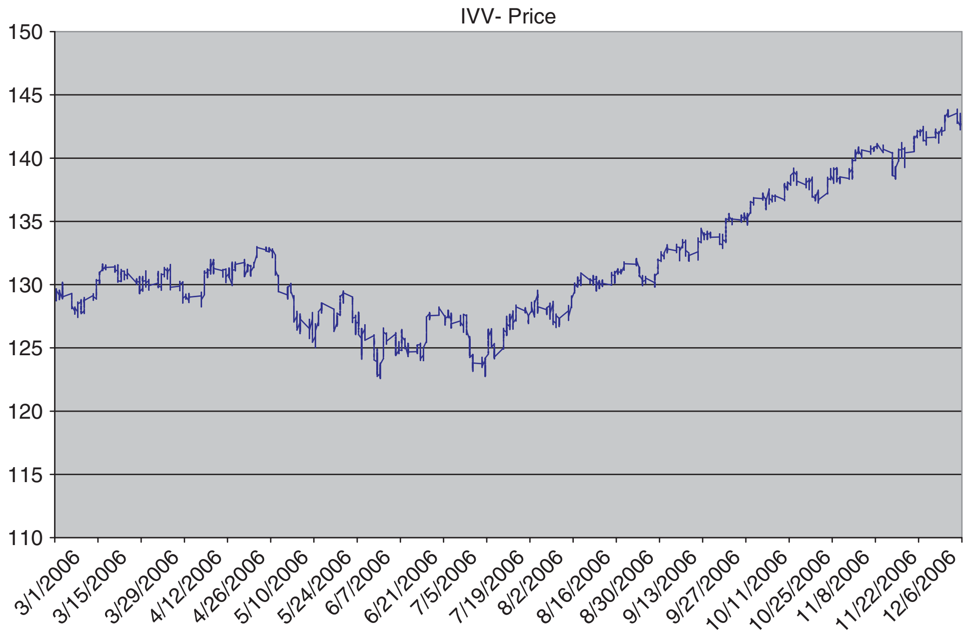


**Fig. 1. IVV price process with unit root**

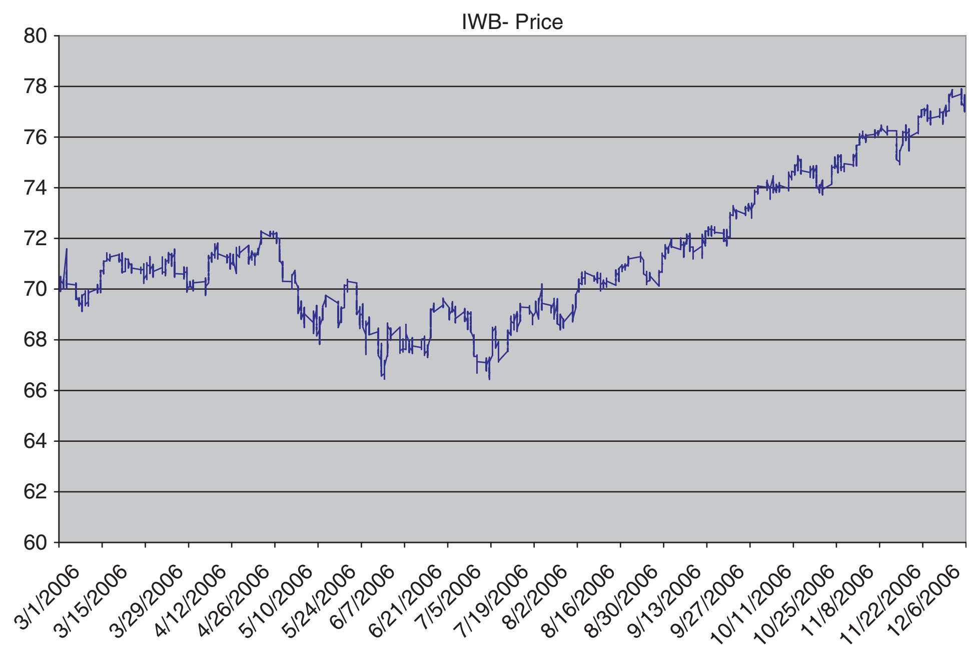


**Fig. 2. IWB price process with unit root**

To develop a strategy that can be implemented it is important to base the strategy only on information that is available as of the time of the transaction. So, to prevent information leakage from the future into the past I determined that the equilibrium relationship between the security prices should be based on the arithmetic average of the ratio of the prices, IVV to IWB. Figure 4 shows the price ratio and the running average of the price ratios.

The trading strategy is rather simple: if the current ratio of prices is greater than the running average price ratio, IVV is overpriced relative to IWB, so IWB should be bought and IVV should be sold short. Specifically, one share of IWB is bought which IWB/IVV shares of IVV are sold short to completely finance the purchase of IWB. If the current price ratio is lower than the running average price ratio, then IWB is

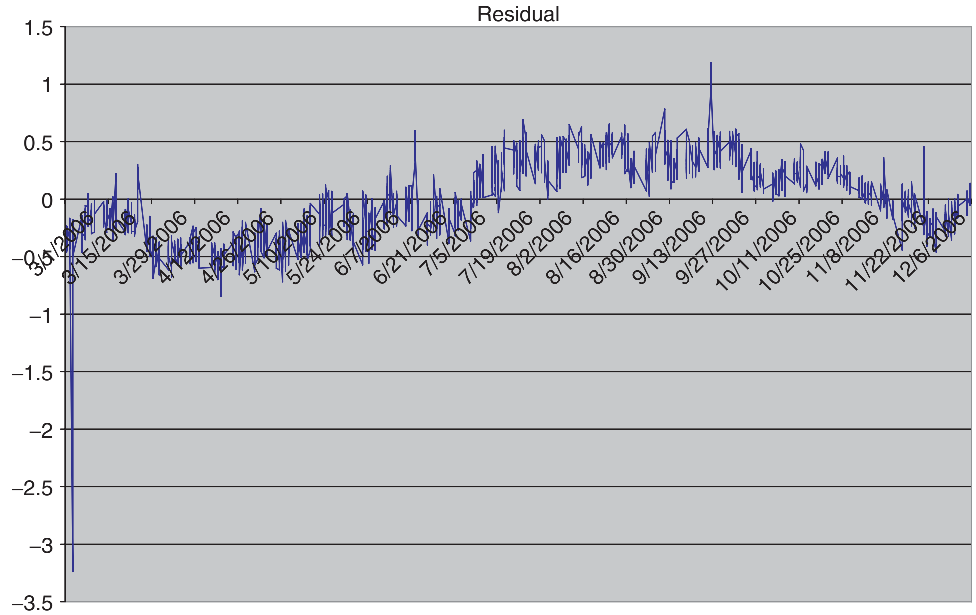


**Fig. 3. Residual process, no unit root**

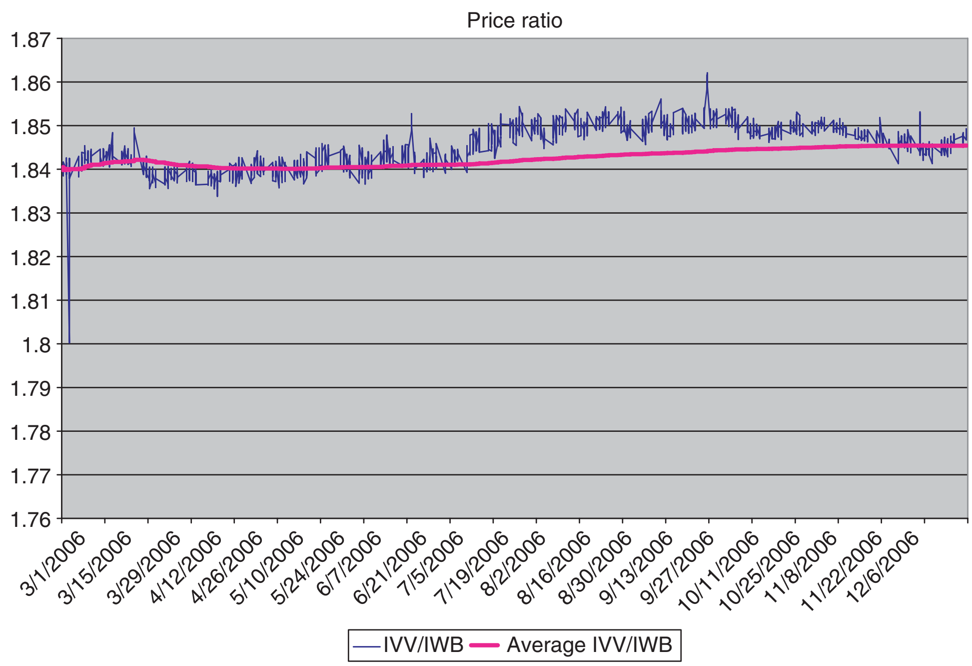


**Fig. 4. Price ratio of IVV to IWB**

overpriced relative to IVV, so one share of IWB is sold short while IWB/IVV shares of IVV are bought.

To make this strategy realistic I assumed a per share transaction cost of \$0.006. To keep transaction costs low, the portfolio was not rebalanced to what the theoretical price ratios should be every time the prices changed. Instead, the portfolio positions were adjusted only according to the following rules: (1) no positions were held over night, so the positions were closed at the last trade of every day and reopened at the theoretical levels every morning; (2) portfolio positions were closed when the overpriced fund switched from one to the other and the positions were re-established at the theoretically optimal levels.

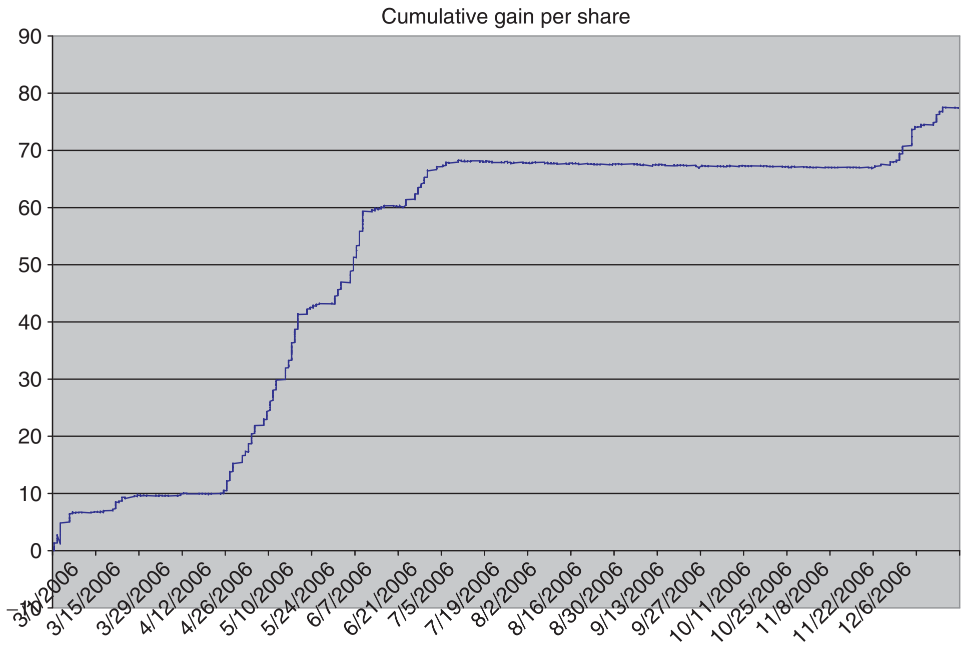


**Fig. 5. Cumulative gain per share**

For a concrete example, the trading began at 8:35 AM on 1 March 2006 with the price ratio at 1.8403. The next transaction, at 8:38 AM was at a relative price of 1.83959. Because the current ratio was less than the current running average of 1.83995 (the two relative price points), IWB was overpriced relative to IVV. So, one share of IWB was sold short while 1/1.83959 shares of IVV were bought. At 8:42 AM, the prices moved to a relative price of 1.84054 and the running average relative price ratio was 1.84014. So, now IVV is overpriced relative to IWB. As a result, the previous positions were closed and a new position established of long one share of IWB and short 1/1.84014 shares of IVV. At 8:44 AM, the relative price ratio was 1.84023 with a running average price ratio of 1.84017. Because IVV remained overpriced relative to IWB, no portfolio changes were made. The only things that would precipitate a portfolio change would be if the trading day closed or if IWB became overpriced relative to IVV.

This clearly can be an active trading strategy unless the relative mispricings persist. While the mispricings persist, the portfolio balance can deviate from the theoretical level of zero because the portfolio is not dynamically hedged. During the time period studied there were a total of 2062 portfolio changes, which means the transaction costs are substantial, amounting to $24.74 per share. However, on net, the strategy is profitable as Fig. 5 shows. The total gain per share, after expenses, is $77.46. A simple buy and hold strategy of investing in IVV would have gained $14.18 per share, while IWB would have gained $7.43 per share. With the active trading strategy, margin calls never became an issue because the lowest the portfolio balance went was –$0.275 per share. Because this is a self-funding strategy, the implementation should be fairly easy provided the infrastructure is in place to monitor the trades.

This article has shown how cointegration and error-correction models can be used in statistical arbitrage applications. Naïve strategies based on simple regression analysis ignore the possibility that the estimated relationship might be spurious due to having nonstationary variables. Cointegration analysis constructs a relationship between the nonstationary variables such that the error term is stationary. This allows for modelling the error correction which can be exploited as a trading strategy.

## References

Alexander, C. and Dimitriu, A. (2005) Indexing and statistical arbitrage, *Journal of Portfolio Management*, **31**, 50–3.

Bondarenko, O. (2003) Statistical arbitrage and securities prices, *The Review of Financial Studies*, **16**, 875–919.

Canjels, E., Prakash-Canjels, G. and Taylor, A. (2004) Measuring market integration: foreign exchange arbitrage and the gold standard, 1879–1913, *The Review of Economics and Statistics*, **86**, 868–82.

Dickey, D. A. and Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 423–31.

Dickey, D., Jansen, D. and Thornton, D. (1991) A primer on cointegration with an application to money and income, *Federal Reserve Bank of St. Louis*, **March**, 58–78.

Forbes, C., Kalb, G. and Kofman, P. (1999) Bayesian arbitrage threshold analysis, *Journal of Business & Economic Statistics*, **17**, 364–72.

Harasty, H. and Roulet, J. (2000) Modeling stock market returns, *Journal of Portfolio Management*, **26**, 33–46.

Hogan, S., Jarrow, R., Teo, M. and Warachka, M. (2004) Testing market efficiency using statistical arbitage with applications to momentum and value trading strategies, *Journal of Financial Economics*, **73**, 525–65.

Kumar, P. and Seppi, D. (1994) Information and index arbitrage, *The Journal of Business*, **67**, 481–509.

Laopodis, N. and Sawhney, B. (2002) Dynamic interactions between Main Street and Wall Street, *The Quarterly Review of Economics and Finance*, **42**, 803–15.

Tatom, J. (2002) Stock prices, inflation and monetary policy, *Business Economics*, **37**, 7–19.

Wang, G. and Yau, J. (Juned 1994) A time series approach to testing for market linkage: unit root and cointegration tests, *The Journal of Futures Markets*, **14**, 457–74.