Basket trading under co-integration with the logistic mixture autoregressive model

Xixin Cheng, Philip L.H. Yu & W.K. Li

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong

Published online: 13 Dec 2010.

To cite this article: Xixin Cheng, Philip L.H. Yu & W.K. Li (2011) Basket trading under co-integration with the logistic mixture autoregressive model, Quantitative Finance, 11:9, 1407-1419, DOI: 10.1080/14697688.2010.506445

To link to this article: http://dx.doi.org/10.1080/14697688.2010.506445

This article is made up of 13 pages.
Basket trading under co-integration with the logistic mixture autoregressive model

XIXIN CHENG*, PHILIP L.H. YU and W.K. LI
Department of Statistics and Actuarial Science,
The University of Hong Kong, Hong Kong
(Received 26 May 2008; revised 29 June 2010; in final form 1 July 2010)

In this paper, we propose a co-integration model with a logistic mixture auto-regressive equilibrium error (co-integrated LMAR), in which the equilibrium relationship among cumulative returns of different financial assets is modelled by a logistic mixture autoregressive time series model. The traditional autoregression (AR) based unit root test (ADF test), used in testing co-integration, cannot give a sound explanation when a time series passes the ADF test. However, its largest root in the AR polynomial is extremely close to, but less than, one, which is most likely the result of a mixture of random-walk and mean-reverting processes in the time series data. With this background, we put an LMAR model into the co-integration framework to identify baskets that have a large spread but are still well co-integrated. A sufficient condition for the stationarity of the LMAR model is given and proved using a Markovian approach. A two-step estimating procedure, combining least-squares estimation and the Expectation-Maximization (EM) algorithm, is given. The Bayesian information criterion (BIC) is used in model selection. The co-integrated LMAR model is applied to basket trading, which is a widely used tool for arbitrage. We use simulation to assess the model in basket trading strategies with the statistical arbitrage feature in equity markets. Data from several sectors of the Hong Kong Hang Seng Index are used in a simulation study on basket trading. Empirical results show that a portfolio using the co-integrated LMAR model has a higher return than portfolios selected by traditional methods. Although the volatility in the return increases, the Sharpe ratio also increases in most cases. This risk–return profile can be explained by the shorter converging period in the co-integrated LMAR model and the larger volatility in the ‘mean-reverting’ regime.

Keywords: Basket trading; Co-integration; Logistic mixture; EM algorithm; Relative value trading

1. Introduction

Basket trading refers to trading a basket of multiple securities simultaneously. It can be traced to the beginning of the 1960s, when the use of computers in trading significantly increased brokerage efficiency, and institutional investors became the major players in the market. Later, when the stock market index and the exchange traded funds (ETFs) became more popular, basket trading became the standard tool for trading the index and the ETFs. About 20 years ago, quantitative trading emerged in the market, with great success in quantitative hedge funds such as Renaissance Technologies and D.E. Shaw. Since then, speculation strategies in the ‘statistical arbitrage’ family have become widely used in the investment world. The framework of statistical arbitrage is to construct a basket of assets that has a statistical stationary net value. Thus, traders can identify whether a basket is significantly overvalued (or undervalued), and make profit by taking advantage of the foreseeable mean reversion. Because the securities in the basket usually need to be traded simultaneously, basket trading becomes the ideal tool to conduct statistical arbitrage strategies. The two most popular statistical arbitrage strategies are pairs trading and ETF arbitrage. In pairs trading, the basket includes two assets in opposite positions, while in the ETF arbitrage, the portfolio includes the ETF and the synthetic ETF in opposite positions.

Currently, three major classes of models are used to drive statistical arbitrage: neural network models, stochastic mean-reverting models and time series
co-integration models. Burgess (1996) considered the typical neural network family. However, due to their over-fitting nature, neural network models require rebalancing of the basket at a very high frequency, probably daily. This leads to extremely high transaction costs, which limit the use of neural network models in broader asset classes. With respect to stochastic mean-reverting models, Anteneodo and Riera (2005) proposed an underlying additive-multiplicative stochastic model for the mean-reverting process. Based on this, Elliott et al. (2005) proposed an unobserved spread model for pairs trading. The disadvantage of the stochastic spread model is that it cannot give guidance in constructing a portfolio; it is rather a model for detecting trading opportunities. Elliott et al. (2005) only applied the model to pairs constructed in the simplest way, that is, those pairs that put an equal dollar amount into both the long and short position. However, this is not always done in practice. Besides, the stochastic spread models do not guarantee profit even if the pair is properly traded according to the open and close signals generated by the model. Also, Elliott et al. (2005) did not include any empirical results in the paper.

The time series co-integration models have their theoretical foundations in Engle and Granger (1987). The co-integration relationship can be described as follows.

The components of the vector time series \( X_t \) are said to be co-integrated of the order \( d, b \) \((d \geq b)\), denoted \( X_t \sim C(\alpha, \beta) \), if (i) all components of \( X_t \) are \( I(d) \); and (ii) a vector \( \alpha (\neq 0) \) exists so that \( X_t \alpha = z_t \sim I(d-b), b > 0 \). The vector \( \alpha \) is called the co-integrating vector and \( z_t \) is called the equilibrium error.

To apply this to the real market, consider the case where each component in \( X_t \) is the cumulative return series of an asset, and \( \alpha \) is the dollar position on each asset. Then \( z_t \) is the net value of the basket. If \( X_t \) follows geometric Brownian motion, then the components in \( X_t \) are \( I(1) \). To be co-integrated, \( z_t \) should be \( I(0) \), that is, it should be stationary. However, most existing research only uses the ARMA model to describe the \( z_t \) series (Dempster et al. 1977, Vidyamurthy 2004, Gatev et al. 2006). As co-integration between asset cumulative returns is a long-term relationship, the stationarity of the \( z_t \) series is more likely due to a mixture of random walks and mean-reverting processes. This results in an ARMA model having the largest root of the AR polynomial being extremely close to one, which makes it hard to identify whether \( z_t \) is a mixture of random walks and fast mean-reverting processes, or whether \( z_t \) is a very slow mean-reverting process.

From another point of view, although Dunis and Ho (2005) showed that there was about 10.4% annualized return for pairs trading from 1999 to 2002, the return under consideration was on the ‘invested capital’, which excluded capital that was not put into trading. The actual monthly returns reported in the paper show that, since 1994, there are more and more negative returns, which are probably due to the simple method used by Dunis and Ho (2005) becoming too ‘crowded’. In this paper, we investigate those baskets that have a large spread due to the non-mean-reverting period, but are still reverting to the mean fast enough. In other words, we try to identify baskets with a stationary net value that is a mixture of a random walk and a mean-reverting process.

To pursue this goal, we combine the co-integration model and the logistic mixture autoregressive model (LMAR), which is studied by Wong and Li (2001a). Logistic mixture time series models were also considered by Jeffries (1998). The LMAR model is a regime-switching time series model that allows different mean structures and volatility levels in different regimes. Also, the probabilities of being in different regimes change dynamically following a logistic linking function.

Section 2 describes the co-integrated LMAR model in detail. To satisfy the co-integration requirement, conditions for the stationarity of the LMAR process are given and proved by embedding the LMAR model into a Markov chain. Section 3 gives the estimation and model selection procedure for the co-integrated LMAR model. Simulation studies for assessing the performance of estimation and model selection are also included in section 3. Following the path of Engle and Granger (1987), a two-step LSE-EM estimation procedure is used to achieve a balance between estimating performance and computational efficiency. The Bayesian information criterion (BIC) is used for model selection. Section 4 derives an application of the co-integrated LMAR model to basket trading. Simulated trading using the co-integrated LMAR model, comparing it with the regression-ADF-test model, is conducted in several sectors and composite stocks of the Hong Kong Hang Seng Index. Empirical results are also reported in this section.

2. The co-integrated LMAR model

Let \( \mathcal{F}_t \) be the \( \sigma \)-field generated by \( \{a_0, a_1, \ldots, a_N\} \). \( F(a_t | \mathcal{F}_{t-1}) \) the conditional cumulative distribution function of \( a_t \) given \( \mathcal{F}_{t-1} \), and \( \Phi(\cdot) \) the cumulative distribution function of the standard Gaussian distribution. The co-integrated LMAR model (co-integrated LMAR) is given as

\[
X_t'\alpha = a_0 + a_t,
\]

where

\[
X_t = (X_{1,t}, X_{2,t}, \ldots, X_{N,t})', \quad \alpha = (a_1, a_2, \ldots, a_N)',
\]

\[
F(a_t | \mathcal{F}_{t-1}) = \sum_{k=1}^{2} p_k, \Phi \left( \frac{e_{k,t}}{\sigma_k} \right),\]

\[
e_{k,t} = a_t - \psi_{k,0} - \sum_{i=1}^{m_k} \psi_{k,i} a_{t-i}, \quad k = 1, 2,
\]

\[\]

\( \dagger \)Following Engle and Granger (1987), a time series with no deterministic component that has a stationary, invertible, ARMA representation after differencing \( d \) times is said to be integrated of order \( d \), denoted \( X_t \sim I(d) \).
and
\[ r_t = \log(p_{1,t}/p_{2,t}) = \delta_0 + \sum_{i=1}^{n} \delta_i |a_{i-1}|, \]
where \( p_{i,t}, i = 1,2 \), are the mixing probabilities, \( p_{1,t} + p_{2,t} = 1 \) for all \( t \), and \( \alpha_0 \) centralizes \( a_t \) to a zero mean process. The equilibrium error \( a_t \) models the deviation spread of the net portfolio value from its mean level, \( \alpha_0 \). The fixed \( \alpha \) (because we want to save the transaction cost) models the dollar position on each asset \( X_{i,t} \). Examples of constructing the associated portfolio will be illustrated in the application section.

The \( a_t \) in (1) follows an LMAR process, where the innovation at time \( t \) given \( F_{t-1} \) can be treated as
\[
\begin{align*}
\sigma_1 w_t, & \quad \text{with probability } p_{1,t}, \\
\sigma_2 w_t, & \quad \text{with probability } p_{2,t}.
\end{align*}
\]
Here, \( w_t \) is a standard white noise process with a zero mean and a variance of one. Thus, for each time \( t, a_t \) can be generated by one AR process in regime 1 with probability \( p_{1,t} \), or it can be generated by another AR process in regime 2 with probability \( p_{2,t} \). We can see that, as \( r_t \) goes towards positive infinity, \( p_{1,t} \) is becoming larger, and as \( r_t \) goes towards negative infinity, \( p_{2,t} \) is becoming larger. Thus, (2) indicates that the LMAR model is actually a dynamic regime-switching model.

In the above model, both the innovation of the \( a_t \) process and its variance are regime dependent. The different values of \( \sigma_i^2 \) take into account changes in the volatility to some extent but, in principle, the innovation series can also have a regime-dependent conditional autoregressive heteroscedastic structure that can be described by the ARCH or GARCH processes (Engle 1982). Incorporating a GARCH structure for \( e_{k,t} \) can be done in principle as in Wong and Li (2001b) and should provide some refinement to the result of this paper. However, since the observed variance is a dynamic mixture of \( \sigma_1^2 \) and \( \sigma_2^2 \), this gives the model some ability to capture the conditional variance pattern. Also, the derivation of theorem 2.1, which is the backbone of this paper, may become much more complicated if a GARCH structure is imposed. We will therefore leave this to a future paper.

Note that model (1) is given in the form of an equilibrium equation, not in the usual error correction form with the vector autoregressive model (VAR) as in other works, for example Maddala and Kim (1998), with the vector equilibrium correction model (VEC) as in Bec and Rahbek (2004), or with co-integration with the GARCH variance model as in Wong et al. (2005).

The reason for structuring the model in this way has its roots in trading practice. As mentioned previously, the model should give a ‘rule’ to construct the basket, which is given by the estimate of the co-integrating vector. However, models in the VAR form, because of their matrix coefficients, have many more parameters than the simple equilibrium equation form, and the co-integration vector must thus be further calculated as the eigenvectors of the structural matrix. All of these estimations raise the computational load and require a large sample size. When facing multiple asset baskets and limited market data, models in the VAR form are not preferred.

For \( X_{i,t} \) to be co-integrated, \( a_t \) should be a stationary process. However, although the simulation study of Wong and Li (2001a) indicates that the presence of one non-stationary regime is possible, the absence of a stationarity result leaves open the question of the conditions that make the whole process globally stationary.

To derive the stationary conditions, note that the use of the LMAR model provides a better description of the spread series \( a_t \). For the purpose of arbitrage, the regime-switching nature identifies those baskets that occasionally have large spreads but are still well co-integrated. These baskets will most likely have larger profits in arbitrage trading. In (1), the absolute spread, \( |a_{i-1}| \), controls the probability of being in a particular regime. It is assumed that regime 1 is stationary, which is used to model the basket as mean-reverting, and regime 2 is non-stationary or even explosive, which is used to model the process that generates the large spread. An intuitive line of thinking is that, if the spread is large enough, arbitrage opportunities would be observed, and arbitrage trading makes the large spread become smaller, which means that, when the realized spread is large enough, the process should almost surely fall back into the stationary regime. This leads to the following theorem.

**Theorem 2.1:** Suppose regime 1 is stationary, regime 2 is non-stationary, and \( \mathbf{a} = (a_{-1}, a_{-2}, \ldots, a_{-n}) \). If \( r_t \to +\infty \) as \( \mathbf{a} \to \infty \), then the process \( a_t \) defined in (1) is stationary. Here, \( \| \cdot \| \) can be defined as either the absolute sum or the square root of the sum of squared components in \( \mathbf{a} \).

The proof of this theorem is presented in the appendix. Generally, it includes two parts. The first is to embed the LMAR model into a Markov chain. In the second part, theorem 15.0.1(iii) of Meyn and Tweedie (1993) is applied to the embedded Markov chain with the conditions stated above, and the ergodicity of the Markov chain can be proved. The existence of higher-order stationarity is also guaranteed. Bec and Rahbek (2004) and Bec et al. (2008) adopted the same approach in their mixture models.

One choice to satisfy the above conditions is that \( \delta_i \) (\( i = 1, \ldots, n \)) are all non-negative with at least one \( \delta_i \) positive. We will use this condition as one of the selection criteria in the application section.

### 3. Model estimation

Since we are given the cumulative distribution function, an obvious estimation method would be maximum likelihood estimation (MLE). MLE is the theoretically preferred method, and it can usually achieve efficiency. However, as mentioned by Engle and Granger (1987), the MLE approach for co-integration models is very time consuming. Thus it is not a preferred method in practice.

In the later application section, we will see why the
estimation time is critical in practice. Hence, we follow the two-step estimation approach of Engle and Granger (1987).

### 3.1. Parameter estimation

The first step in the estimation procedure is to estimate the co-integration vector $\alpha$ and intercept $\alpha_0$. Since we have proved the stationarity of the LMAR model followed by $a_t$ under the conditions stated in theorem 2.1, it is clear that the co-integration relationship between $\{X_{i,t}\}$ is well defined. We use 'co-integrating regression', which is essentially a least-squares estimation (LSE) method:

$$
(a_0, \hat{\alpha}) = \arg \min_{a_0, \alpha} \| X_t' \alpha - a_0 \|^2. \tag{3}
$$

Of course, some normalization must be imposed on $\alpha$. In this model, as $\alpha$ is the 'rule' to construct the portfolio, we choose one $X_{1,t}$ in $\{X_{i,t}\}$ as the response variable and regress the selected $X_{1,t}$ on the other $X_{i,t}$. The choice of the response $X_{1,t}$ depends on the context of each case. In pairs trading, we choose $X_{1,t}$ as the response variable. In ETF arbitrage trading, clearly we choose ETF as the response variable.

The property of this estimator is discussed in theorem 2 of Engle and Granger (1987) in great detail. This estimator has the same limiting distribution as the MLE estimator. Least-squares standard errors will be consistent estimates of the true standard errors. Although the choice of which variable is the response variable does have some impact on the value of $\hat{\alpha}$, this impact matters very little as the sample size increases. This conclusion is also confirmed by our simulation studies. Moreover, this estimator is generally independent from the parameter estimators in the LMAR model.

After obtaining $\hat{\alpha}$, parameters in the LMAR model are estimated with the Expectation-Maximization (EM) algorithm. The EM algorithm was first proposed by Dempster et al. (1977), and is used as a tool to find the MLE when the true likelihood function is hard or even impossible to obtain or to optimize. Ruud (1991) reviewed extensions of the EM algorithm. The MLE nature of the EM algorithm makes it a powerful tool to perform MLE estimation for mixture models. Wong and Li (2001a,b) and Bec and Rahbek (2004) showed that the EM algorithm has good performance in estimating mixture models.

Let $Z_{k,t}$ be the indicator variable of regime $k$ ($k = 1, 2$), where $Z_{1,t} + Z_{2,t} = 1$ for every $t$. Define the following:

$$
Y = (a_1, a_2, \ldots, a_N), \quad \xi = (\delta_0, \delta_1, \ldots, \delta_n),
$$

$$
\zeta_1 = (\varphi_1, 0, \varphi_1, \ldots, \varphi_1, m_1), \quad \zeta_2 = (\varphi_2, 0, \varphi_2, \ldots, \varphi_2, m_2), \quad \Theta = (\xi, \zeta_1, \zeta_2, \sigma_1^2, \sigma_2^2), \quad s = \max(m_1, m_2, n).
$$

In practice, $Z_1$ and $Z_2$ are unknown. To proceed, we first consider them as if they were known and call $(Y, Z_1, Z_2)$ the complete data, following standard statistical terminology (Tanner 1993). $Z_1$ and $Z_2$ will be estimated in the EM algorithm below. The log-likelihood for the complete data $(Y, Z_1, Z_2)$, conditional on $\Theta$, is given by

$$
I = \sum_{t=1}^{N} \sum_{k=1}^{2} Z_{k,t} \left( \log(p_{k,t}) - \frac{1}{2} \log(\sigma_k^2) - \frac{\varepsilon_{k,t}^2}{2\sigma_k^2} \right). \tag{4}
$$

In the E step, we have $\hat{\Theta}^{(i-1)}$, which is the parameter estimation in the $(i-1)$ iteration. We need to compute $E(I \mid Y, \hat{\Theta}^{(i-1)})$. Denote $\Phi(x)$ by $\phi(x)$. In our case, we just replace $Z_{k,t}$ in (4) with $\hat{Z}_{k,t}^{(i)} = E(Z_{k,t} \mid Y, \hat{\Theta}^{(i-1)})$, where

$$
\hat{\varepsilon}_{k,t}^{(i)} = \frac{p_{k,t}}{\sigma_k} \phi \left( \frac{\hat{c}_{k,t}}{\sigma_k} \right) \left[ \frac{p_{1,t}}{\sigma_1} \phi \left( \frac{\hat{c}_{1,t}}{\sigma_1} \right) + \frac{p_{2,t}}{\sigma_2} \phi \left( \frac{\hat{c}_{2,t}}{\sigma_2} \right) \right]_{\hat{\Theta}^{(i-1)}}.
$$

After replacing $Z_{k,t}$ with $\hat{Z}_{k,t}^{(i)}$ in (4), we denote this new function as $Q^{(i)}(\Theta)$.

In the M step, we estimate the parameter in the $i$th step by the optimization

$$
\hat{\Theta}^{(i)} = \arg \max_{\Theta} Q^{(i)}(\Theta).
$$

Note that the first-order derivatives of $Q$ with respect to $\Theta$ are

$$
\frac{\partial Q}{\partial \varepsilon} = \sum_i (\varepsilon_{1,t} - p_{1,t}) \frac{\partial r_i}{\partial \varepsilon}, \quad \frac{\partial Q}{\partial \zeta_k} = -\sum_i t_{k,i} \frac{\partial \xi_k}{\partial \varepsilon}, \quad k = 1, 2,
$$

and

$$
\frac{\partial Q}{\partial \sigma_k} = \sum_i t_{k,i} \left( \frac{\varepsilon_{k,t}^2}{\sigma_k^2} - 1 \right), \quad k = 1, 2,
$$

where

$$
\frac{\partial r_i}{\partial \varepsilon} = (1, [a_{i-1}, 1, a_{i-2}], \ldots, \varepsilon_{i-1}^T), \quad \frac{\partial \xi_k}{\partial \varepsilon} = -\left(1, a_{i-1}, a_{i-2}, \ldots, a_{i-m} \right)^T.
$$

The second-order derivatives of $Q$ with respect to $\Theta$ are

$$
\frac{\partial^2 Q}{\partial \varepsilon^2} = \sum_i p_{1,i} (1 - p_{1,i}) \frac{\partial r_i}{\partial \varepsilon} \frac{\partial r_i}{\partial \varepsilon}, \quad \frac{\partial^2 Q}{\partial \zeta_k^2} = -\sum_i t_{k,i} \frac{\partial \xi_k}{\partial \varepsilon} \frac{\partial \xi_k}{\partial \varepsilon}, \quad k = 1, 2,
$$

and

$$
\frac{\partial^2 Q}{\partial \sigma_k^2} = \sum_i t_{k,i} \left( \frac{\varepsilon_{k,t}^2}{\sigma_k^2} - 1 - \frac{\sigma_k^2}{\sigma_k^2} \right), \quad k = 1, 2,
$$

and

$$
\frac{\partial^2 Q}{\partial \sigma_1 \sigma_2} = \sum_i t_{k,i} \left( \frac{\varepsilon_{k,t}^2}{\sigma_1^2} - 1 - \frac{\sigma_1^2}{\sigma_2^2} \right), \quad k = 1, 2.
$$

Because $\hat{\Theta}^{(i)}$ has no analytical form, we use the Newton–Raphson method. Given $\hat{\xi}^{(i-1)}$, we take

$$
\hat{\xi}^{(i)} = \hat{\xi}^{(i-1)} - \left( \frac{\partial^2 Q}{\partial \xi^2} \right)^{-1} \frac{\partial Q}{\partial \xi}.
$$

The $\hat{\Theta}^{(i)}$ is determined by iterating the above equation until convergence.
The \( i \)th step estimators for \( \zeta_k \) and \( \sigma_k^2 \) fall into the framework of generalized least-square (GLS) estimation, and they have analytical forms, for \( k = 1, 2 \),

\[
\hat{\zeta}^{(i)}_k = \left( \sum_i \tau_{k,i} \frac{\partial \psi_{k,i} \partial \psi_{k,i}}{\partial \psi_{k,i}} \right)^{-1} \left( \sum_i \tau_{k,i} \left( a_i + \frac{\partial \psi_{k,i} \partial \psi_{k,i}}{\partial \psi_{k,i}} \right) \right),
\]

\[
\hat{\sigma}^{(i)}_k = \left( \sum_i \tau_{k,i} \right)^{-1} \left( \sum_i \tau_{k,i} \left( a_i + \frac{\partial \psi_{k,i} \partial \psi_{k,i}}{\partial \psi_{k,i}} \right) \right).
\]

The EM algorithm stops when \( \| \hat{\Theta}^{(i)} - \hat{\Theta}^{(i-1)} \| \) is small enough to be considered as being converged.

The information matrix obtained by the EM algorithm is different from the information matrix of the normal MLE estimator. Louis (1982) gave the following formula for calculating the observed information matrix obtained by the EM algorithm:

\[
I = I_{\text{complete}} - I_{\text{missing}} = E \left( -\frac{\partial^2 l}{\partial \Theta^2} \mid Y, \Theta \right) \| \Theta \_0
\]

\[
- \text{var} \left( \frac{\partial l}{\partial \Theta} \mid Y, \Theta \right) \| \Theta \_0.
\]

\( E(-\frac{\partial^2 l}{\partial \Theta^2} \mid Y, \Theta \) has been given in our case as the Hessian matrix. The \( \text{var}(\frac{\partial l}{\partial \Theta} \mid Y, \Theta) \) is given as

\[
\sum_i p_{i,1}(1 - p_{i,1})V_i^2,
\]

where

\[
V_i = \frac{\partial \psi_{i,1} \partial \psi_{i,1}}{\sigma_i^2} - \frac{e_{i,1} \partial \psi_{i,1}}{\sigma_i^2} \left( \frac{e_{i,1}}{\sigma_i^2} - 1 \right),
\]

\[
- \frac{e_{i,2} \partial \psi_{i,2}}{\sigma_i^2} \left( \frac{e_{i,2}}{\sigma_i^2} - 1 \right).
\]

The asymptotic variance of the parameters in the LMAR part is obtained by the diagonal elements in the inverse of the observed information matrix \( I \).

### 3.2. Simulation studies and model selection

To assess the estimation performance of the two-step estimation procedure, we conducted various simulation studies. In all the simulation studies, \( X_{1,t} \) was used as the response variable in the regression. The simulation study we report here is based on the following model:

\[
X_{1,t} = 0.3X_{2,t} - 0.8X_{3,t} + \alpha_t - 2,
\]

where

\[
X_{2,t} = X_{2,t-1} + 2u_t,
\]

\[
X_{3,t} = X_{3,t-1} + 4v_t,
\]

\[
F(a_t \mid F_{t-1}) = p_{i,1} \Phi \left( \frac{e_{i,1}}{\sigma_i} \right) + p_{i,2} \Phi \left( \frac{e_{i,2}}{\sigma_i} \right),
\]

\[
e_{i,1} = a_t - (0.6a_{t-1} - 0.2a_{t-2}),
\]

\[
e_{i,2} = a_t - 1.5a_{t-1},
\]

\[
r_t = \log p_{1,t}/p_{2,t} = -1.3 + 0.6|a_{t-1}| + 0.3|a_{t-2}|.
\]

In the above model, \( u_t \) and \( v_t \) are independent standard Gaussian white noise processes. The \( X_{2,t} \) and \( X_{3,t} \) are aimed at simulating the Brownian motion in cumulative returns of the assets with different volatilities. The time series structure in \( e_{1,t} \) represents the stationary regime under normal market conditions. The time series structure in \( e_{2,t} \) represents the explosive regime in the chaotic market period. In \( a_t \), a mixture of volatility is also adopted to reflect the different volatilities in different market periods. Also, note that \( \delta_0 = -1.3 \), which means that the probability of being in the stationary regime has a lower bound of 21.42\% \((= e^{-1.3}/(1 + e^{-1.3}))\). Although it looks as if \( a_t \) has a greater probability of being in a non-stationary region, stationarity is still guaranteed by theorem 2.1. The simulation runs for 300 replications, and each replication has a sample size of 500.

The results of the simulation study are reported in table 1. For the co-integrating part, LSE gives a good estimation for \( \alpha \) and \( \alpha_0 \). However, due to the time-dependent structure in \( a_t \), the standard error of the LSE estimator is under-estimated. Note that the empirical standard errors for \( \alpha \) and \( \alpha_0 \) are mostly two times the theoretical standard errors. This can be explained by the reduction of the effective sample size in \( a_t \) due to the regime-switching AR structure. The actual effective sample size in each regime is only about half of the sample size, but the formulae of standard errors in LSE still use the original sample size as the divisor. Figure 1 shows the partial autocorrelation function of one piece of simulated data. The partial autocorrelation function is an effective tool for selecting the embedding dimension (Kantz and Schreiber 2001). We can see from figure 1 that the embedding dimension is between two and three. Except for the co-integrated parameters, the average theoretical standard errors and the empirical standard errors generally match each other well. For the co-integrated parameters, what is reassuring is that, although the estimated theoretical standard error has been under-estimated, our overall judgment on the significance of \( \alpha \) and \( \alpha_0 \) is unaffected if we multiply the estimated theoretical standard errors by a factor of 5 or even more. In addition, the LSE procedure saves substantial estimating time, which is another advantage that we will see in the application section.

For the parameters in the time series structure, the EM algorithm gives a satisfactory estimation result, with a small bias and an acceptable standard error. Clearly, all the parameter estimates are significant. The relatively large bias for parameters in the logistic mixture component is because the complete-data log-likelihood function, i.e. the \( Q \) function, is less sensitive to parameters in the logistic mixture component.

With respect to model selection, we consider the Bayesian Information Criterion (BIC) (Schwarz 1978) as the model selection method. For our model, the BIC formula is

\[
\text{BIC} = -2l^* + \log(N - \max(m_1, m_2, n))(3 + m_1 + m_2 + n),
\]

where \( N \) is the sample size and \( l^* \) is the maximized log-likelihood function for the observed data, using parameters estimated from the EM algorithm given \( m_1 \), \( m_2 \) and \( n \). The chosen model dimension should be associated with the smallest BIC value.
Table 1. Estimation results of the simulation study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Average of estimates</th>
<th>Average estimated theoretical SE</th>
<th>Empirical SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-integrating vector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>$-2$</td>
<td>$2.0384$</td>
<td>$0.3353$</td>
<td>$0.6836$</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$-0.3$</td>
<td>$-0.2996$</td>
<td>$0.0055$</td>
<td>$0.0108$</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$-0.8$</td>
<td>$-0.8009$</td>
<td>$0.0181$</td>
<td>$0.0360$</td>
</tr>
<tr>
<td>Regime 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>$1$</td>
<td>$0.9955$</td>
<td>$0.1411$</td>
<td>$0.1545$</td>
</tr>
<tr>
<td>$\phi_{1,0}$</td>
<td>$0$</td>
<td>$-0.0173$</td>
<td>$0.0795$</td>
<td>$0.1122$</td>
</tr>
<tr>
<td>$\phi_{1,1}$</td>
<td>$0.6$</td>
<td>$0.5937$</td>
<td>$0.0368$</td>
<td>$0.0392$</td>
</tr>
<tr>
<td>$\phi_{1,2}$</td>
<td>$-0.2$</td>
<td>$-0.1987$</td>
<td>$0.0393$</td>
<td>$0.0405$</td>
</tr>
<tr>
<td>Regime 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>$3$</td>
<td>$3.0098$</td>
<td>$0.3498$</td>
<td>$0.3462$</td>
</tr>
<tr>
<td>$\phi_{2,0}$</td>
<td>$0$</td>
<td>$0.0215$</td>
<td>$0.1409$</td>
<td>$0.1883$</td>
</tr>
<tr>
<td>$\phi_{2,1}$</td>
<td>$1.5$</td>
<td>$1.4933$</td>
<td>$0.1374$</td>
<td>$0.1438$</td>
</tr>
<tr>
<td>Logistic mixture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_0$</td>
<td>$-1.3$</td>
<td>$-1.2924$</td>
<td>$0.5225$</td>
<td>$0.5116$</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>$0.6$</td>
<td>$0.5940$</td>
<td>$0.1612$</td>
<td>$0.1609$</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>$0.3$</td>
<td>$0.3351$</td>
<td>$0.1421$</td>
<td>$0.1585$</td>
</tr>
</tbody>
</table>

Figure 1. Partial autocorrelation function of one simulated realization.

Under the framework of the above model, we ran two batches of replications with a sample size of 500 for the first batch and 800 for the second batch. In each replication, $m_1$, $m_2$ and $n$ are set to an under-fitting value, the true value, and an over-fitting value, respectively ($m_1 = 1, 2, 3, m_2 = 0, 1, 2$ and $n = 1, 2, 3$). With the simulated data, we ran the EM estimation procedure 27 ($= 3^3$) times. Combinations of $\{m_1, m_2, n\}$ with the smallest BIC value are selected. For the first batch of 300 replications with sample size 500 in each replication, 187 replications (62.33%) have the true dimensions identified in all three components. Further, nearly all 300 replications have correctly identified $m_1$ and $m_2$ (293 for $m_1$ and 295 for $m_2$), and 259 replications (86.33%) have correctly identified $n$. These results suggest that the BIC is suitable for selecting the dimension for the model. In most mis-identified cases, $n$ is selected as one instead of the true value of two. This may be because, when $N$ is small, the estimated theoretical standard error is large so that $\delta_2$ is marginally insignificant. However, as sample size increases, the error rate becomes acceptable.

4. Application to basket trading

4.1. Methodology, trading steps and data

The basic trading idea in basket trading is to construct a basket that has a stationary $a_t$. Because $a_t$ is the deviation of the net value from its stationary net value $\delta_0$, the stationarity of $a_t$ means that we can identify whether the portfolio is significantly overvalued (or undervalued), so that we can make profit by shorting (or longing) the portfolio. Vidyamurthy (2004) discussed this statistical arbitrage in great detail, but only the ARMA model was adopted to model the $a_t$ series. In our application, the $a_t$ series generated from the first LSE step is estimated with a LMAR model instead of an AR model. The rules to open and close the basket are controlled by the level of $p_{1,t}$, which is the probability that $a_t$ will fall into the stationary regime. Note that

$$\log(p_{1,t}/(1-p_{1,t})) = r_t = \delta_0 + \sum_{j=1}^{n} \delta_j |a_{t-j}|,$$

where $p_{1,t}$ utilizes the information in $a_{t-1} \ldots a_{t-n}$. This gives the present approach an advantage when compared with the traditional trading rule, which depends solely on $a_t$ itself. Also, this ‘embedded’ method is consistent with the fact that technical traders usually trade based on the prices in the past several time points. In addition,
because $a_t$ has a stationary distribution, so does $|a_t|$. This leads to the result that $p_{1,t}$ should follow some kind of distribution. An intuitive trading rule based on $p_{1,t}$ could thus be as follows: open the basket when $p_{1,t}$ exceeds a certain large percentile; close the basket when $p_{1,t}$ goes back to its median level. Since we expect that the distribution of $p_{1,t}$ is heavily skewed, the quantile approach, rather than the traditional ‘two-standard-error’ approach, is natural. This approach also implies that we should close the basket when $a_t$ is at some level close to zero, rather than waiting for $a_t$ to cross zero in the traditional approach. Because when $a_t$ is close to zero a ‘random walk’ dominates the price pattern, closing the basket when the mean-reverting pattern is not obvious can shorten the open period, which also serves as a means of profit protection.

To be specific, for each trading day, there are two stages called the selection stage and the trading stage. In the selection stage, a training period is given, for each given $N$-asset basket.

1. Construct $X_{1,t}$ as the percentage cumulative return (because the basket log-return is not a linear combination of individual asset log-returns, we do not use a cumulative log-return) series for each asset in the basket within the training period.
2. Estimate parameters in the co-integrated LMAR model with $X_t$.
3. Select the basket in a pool that has a stationary regime 1 with at least one of $\delta_i$ ($i = 1, \ldots, n$) significantly positive, as indicated by two times the estimated standard error, and the other $\delta_i$ insignificantly different from zero. This criterion is motivated by the discussion on the choice of $a_t$ after theorem 2.1.

For a given ‘market’, we go through all possible combinations of $N$-asset baskets. One can expect this to be a huge number. For example, if a market has 20 stocks, then the total number of all four-asset portfolios is $C_4^{20} = 48,454$, and the total number of all five-asset portfolios is $C_5^{20} = 15,504$. From our limited experience the computing time of the LSE-EM approach is more than 10 times faster than the full MLE approach. That is why the LSE-EM estimation procedure is adopted here.

At the end of the selection stage, we have a pool in which baskets are sorted by the standard error of the estimated $a_t$ series in ascending order. The top 10 baskets are referred to as the potential tradable baskets, and are monitored in the next trading stage. In the trading stage, we monitor the $p_{1,t}$ of the potential tradable pair in the pool.

1. Open the basket position according to $\hat{a}$ when $p_{1,t}$ is above its 95% quantile based on the estimated $p_{1,t}$ series in the training period.
2. Set up an observing period, and if the basket is not open within the observing period, it is removed from the pool, and then needs to be re-estimated.
3. Close the basket position when $p_{1,t}$ crosses its median based on the estimated $p_{1,t}$ series in the training period.

4. Set up a maximum trading life, and if the open basket position is not closed within the trading life, it is unwound on the last day of its trading life.

In our application, we mainly focus on the equity market. We do not consider any margin trading. The transaction fee is included as 0.3% of the total trading amount for both buy and sell. As in the previous simulation study, 500 is a proper size for the training data in terms of efficient estimation and acceptable performance in model selection. Here, we use the most recent 500 trading days as the training window, i.e. approximately two years of historical data of assets. The observing period is set at 25 trading days, i.e. approximately one month. The maximum trading life is set at 50 trading days, i.e. approximately two months. The testing horizon in this application is from 1 January 2007 to 1 January 2008. Thus the data we used started from 500 trading days before 1 January 2007 and went up to 1 January 2008. All the data used in the application are obtained from Yahoo Finance (finance.yahoo.com) with all dividends fully reinvested (the adjusted closing price).

4.2. Trading performance measure and model comparison

Although under the co-integration framework the basket is no longer a completely self-financing basket (the amounts invested in the long leg and the short leg are not the same), the trading performance measure system proposed by Gatev et al. (2006) can still be adopted. We measure the return of a single basket using the return on the required investment, which is the amount invested in the long leg, because normally short selling needs an equal amount of collateralized cash and we need to use proceeds from short selling as the collateralized cash. Each trade is set to have a required investment of $500,000 to avoid any possible rounding error.

Since during the simulation the simulated portfolio can take on a different number of basket positions at one time, we measure the performance of the simulated portfolio with the return on the committed capital. The committed capital is the maximum amount of required investment in the portfolio during the simulation. For example, if during the simulation the maximum number of basket positions that the portfolio could take is four, then the committed capital is $2,000,000 (= 500,000 \times 4)$. The return on committed capital takes the opportunity cost into account, because even if the portfolio only takes on two basket positions, the return will still be diluted. The cumulative profit of the portfolio is marked-to-market daily. At the end of the simulation, the committed capital is decided, and the cumulative return is calculated as the cumulative profit divided by the committed capital.

Using a simulated portfolio, we compare the performance of the co-integrated LMAR model with the traditional model using the regression plus ADF test (RADF). In the traditional approach, the augmented
Dickey–Fuller (ADF) test will be applied to the \( a_t \) at the LSE step. The null hypothesis of the ADF test is that the time series has a unit root. If the test rejects the null hypothesis, the time series is considered to be stationary. Thus, baskets that have \( a_t \) series that reject the null hypothesis in the ADF test are considered as co-integrated baskets, and are put into the pool. To remain consistent, baskets selected into the pool are sorted by the standard error of the estimated \( a_t \) series in ascending order. The top 10 baskets are referred to as the potential tradable baskets, and are monitored. The trading rule is the traditional ‘two-standard-error’ rule: open the basket if \( a_t \) diverges more than two historical standard deviations and close the basket if \( a_t \) crosses zero. The training period, observing period, and maximum trading life in the RADF model are the same as those in the co-integrated LMAR model.

### 4.3. A trading example

Table 2 and figure 2 illustrate one trading example of the two-asset basket, which is also referred to as pairs trading. The basket contains stocks 0363.HK and 0882.HK which are listed on the Hong Kong Stock Exchange. The selection date is 26 September 2007. Thus, the trading period is 500 trading days prior to 26 September 2007. The response variable in the regression is 0363.HK.

The basket contains stocks 0363.HK and 0882.HK which are listed on the Hong Kong Stock Exchange. The two-asset basket, which is also referred to as pairs trading.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>LMAR</th>
<th>AR(_{-N})</th>
<th>AR(_{-T})</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_0 )</td>
<td>5.589(2.7719)</td>
<td>5.589(2.7719)</td>
<td>5.589(2.7719)</td>
</tr>
<tr>
<td>( a_1 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>-0.6884(0.0134)</td>
<td>-0.6884(0.0134)</td>
<td>-0.6884(0.0134)</td>
</tr>
<tr>
<td>( \phi_{1,0} )</td>
<td>73.2200(14.4670)</td>
<td>22.5780(0.6550)</td>
<td>22.5780(0.1692)</td>
</tr>
<tr>
<td>( \phi_{1,1} )</td>
<td>0.2017(0.9101)</td>
<td>0.0239(0.2191)</td>
<td>-0.0406(0.1423)</td>
</tr>
<tr>
<td>( \phi_{2,0} )</td>
<td>0.9110(0.0373)</td>
<td>0.9656(0.0087)</td>
<td>0.9893(0.0077)</td>
</tr>
<tr>
<td>( \phi_{2,1} )</td>
<td>5.9193(0.9056)</td>
<td>0.0052(0.1537)</td>
<td>0.9994(0.0100)</td>
</tr>
<tr>
<td>( \phi_{2,2} )</td>
<td>-2.0516(0.4411)</td>
<td>0.0499(0.0152)</td>
<td>2.7752(0.3641)</td>
</tr>
<tr>
<td>( \delta_0 )</td>
<td>-0.6884(0.0134)</td>
<td>0.9656(0.0087)</td>
<td>0.9893(0.0077)</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>2.7752(0.3641)</td>
<td>0.9656(0.0087)</td>
<td>0.9893(0.0077)</td>
</tr>
<tr>
<td>Log-likelihood value</td>
<td>-1377.2</td>
<td>-1488.71</td>
<td>-1392.9388</td>
</tr>
<tr>
<td>BIC value</td>
<td>2809.7573</td>
<td>2996.0498</td>
<td>2810.7280</td>
</tr>
</tbody>
</table>

Table 2. Estimation results of the trading example.

From table 2 we find that, except for the constant items, all estimators are significant. The two prevalent practice models both give close-to-one roots. Although the spread series \( a_t \) can pass the traditional Dickey–Fuller unit root test (DF test) or its augmented version (the ADF test), their close-to-one roots as well as the corresponding upper boundary of the confidence intervals make this portfolio uncomfortable to trade with. In the estimation results of the LMAR model, there are clearly two regimes. Because one is within the two standard deviation interval of \( \phi_{2,1} \), we can refer to regime 2 as the ‘random-walk’ regime. Regime 1 is clearly stationary; thus we refer to regime 1 as the ‘mean-reverting regime’. The \( \delta_0 \) and \( \delta_1 \) suggest that, although most of the time \( a_t \) falls into the ‘random-walk’ regime, there is a lower bound of 11.39% \((=e^{-2.0516}(1+e^{-2.0516}))\) where \( a_t \) goes into the ‘reverting regime’, so that \( a_t \) is still well co-integrated.

Figure 2 has two panels. The top graph in panel (a) shows the cumulative returns for the two stocks in the training period. As mentioned by Gatev et al. (2006), monitoring the cumulative returns and assessing the average historical deviation between two cumulative returns is the most widely adopted method in real trading practice. However, the graph shows that there is seldom a cross between the two cumulative returns. Thus this pair is usually not traded. The middle graph of panel (a) shows the spread \( a_t \) in the training period. The lower graph of panel (a) shows the probability that \( a_t \) goes into the ‘reverting regime’ in the training period. We observe the pattern that when \( a_t \) is most probably in the ‘random-walk’ regime, it has a lower volatility and when \( a_t \) has a greater probability of being in the ‘mean-reverting regime’, the volatility is greater. This pattern fits the parameter estimation results in table 2. Panel (b) shows the situation in the trading period. The two horizontal lines in the bottom graph of panel (b) indicate the trading boundaries. Hitting the dotted line produces the open signal and hitting the dashed line produces the close signal.

On 26 September 2007, this basket is put into the monitored pool. On the 17th trading day after the selection, \( a_t \) up crosses the dotted line. On 26 October 2007, the 21st trading day after the selection, \( a_t \) down crosses the dotted line, which leads to an open signal (this treatment is a risk management method (Lin et al. 2006)). Because \( a_t = (1, -0.6884) \), the basket is defined as follows: for every one dollar long position in 0363.HK, there is a 0.6884 dollar short position in 0882.HK. The positive \( a_t \) means the basket is overvalued,
so we should short the basket. The total out-cashflow is 503,679 dollars, comprising two parts:

- 500,000 dollars for longing 0882.HK; and
- 3679 dollars \( (0.3\% \times \frac{500,000 + 500,000}{0.6884}) \) for the transaction cost. (Remember that no cash flow is received from short-selling.)

On 8 November 2007, \( a_t \) crosses the dashed line, and the short position of the basket is closed. The total in-cashflow is 587,175 \( (=440,750 + 149,477 - 3052) \) dollars comprising three parts:

- 11.85% loss on the long position, and thus a 440,750 dollar in-cashflow from closing the long position in 0882.HK;
- 20.58% profit from the short position, and thus a 149,477 dollar in-cashflow from closing the short position in 0363.HK; and
- 3052 dollars \( (0.3\% \times \frac{440,750 + 500,000}{0.6884} - 1 - 0.2058)) \) for the transaction cost.

The return for this trading is 16.58%. Note that the transaction costs reduce the return by almost 1.5%; this is why we prefer a fixed \( \alpha \) in the model, because portfolio rebalancing incurs large transaction costs, which significantly reduce the return.

### 4.4. Trading results

The simulated trading is conducted in the conglomerate sector, the consumer goods sector, the financial sector and composite stocks in the Hang Seng Index. The stocks in each sector are selected from stock lists of the Hang Seng Composite Index Series and the Hang Seng Index (www.hsi.com.hk), with each stock having at least 3 years of historical data. Due to the large computational load, we only run the two-asset combinations.

Tables 3 and 4 show the empirical results. The average number of days to open is almost the same for both methods. However, the average number of days to close is very different in the co-integrated LMAR model and the RADF model. The average number of days to close in the co-integrated LMAR model is clearly much smaller than that in the RADF model, which demonstrates that the initial design of the co-integrated LMAR model works well in a real-life situation. This quick mean-reverting feature also helps to reduce the committed capital, hence increasing the annual return on committed capital in all four cases, because, the faster one basket is closed, the smaller the number of basket positions the portfolio would take at the same time.

The percentage of open basket positions that are forced to close is also smaller in the co-integrated LMAR model than in the RADF model. This is important for the risk management of basket trading. As long as the basket position is closed before the maximum trading period has expired, the return of single basket trading is always positive (except for some rare cases in which the transaction costs make the return of the single trading negative). Huge losses might occur when the basket positions do not converge as expected, and this may be
interpretation, risk management of the basket trading needs to decide whether to unwind the position or take the position for another period. Simulation also shows that huge losses do happen when basket positions do not converge as expected. The customized exchange options can be used to hedge this kind of risk, but further investigation is needed.

The annual return (on the committed capital) is higher in the co-integrated LMAR model, but the annualized volatility is also higher. To explain this return–risk profile, we recall the basic philosophy of this model. The co-integrated LMAR model tries to identify baskets whose spread $a_t$ is a mixture of ‘random-walk’ and ‘mean-reverting’ processes, and it takes advantage of the fast reverting feature in the mean-reverting regime. From table 2, we find that the mean-reverting regime, although stationary, is associated with a very high volatility. Unfortunately, this is what always happens in the market. Thus, when we open a basket position, we also expose ourselves to a high volatility process, hence increasing the volatility in the return. The bottom line is that although the volatility is increasing, given a reasonable risk-free rate, the Sharpe ratio for the return in the co-integrated LMAR model is usually higher than its counterpart in the RADF model. Note that the profiles for the single trading return in both models are not significantly different. However, because the trading rule in the co-integrated LMAR model is more conservative—we close the basket position before the spread $a_t$ returns to 0—then the deviation in baskets traded by the co-integrated LMAR model is larger, which may also increase the volatility in the return. Taking multiple basket positions at the same time and the increased efficiency in capital utilization also magnify the volatility in the return.

Therefore, if we decompose the return into three sources—the marginal profit rate, the capital efficiency, and the financial leverage—we can see that the marginal profit rate (single trading return) and the financial leverage for the two models are generally the same, and that the co-integrated LMAR model increases the return by increasing the capital efficiency while keeping the marginal profit rate.

5. Conclusion

In this paper, we propose a co-integrated LMAR model, and apply the model to basket trading, especially in the equity market where significant transaction costs would occur. The design of the model tries to shorten the converging period of the basket position, hence increasing the return, by identifying baskets whose spread process $a_t$ is a stationary process that can be regarded as a mixture of ‘random-walk’ and ‘mean-reverting’ processes. The LSE-EM two-step estimating procedure is used to achieve a balance between computational load and estimation performance. The BIC is used for model selection.

### Table 3. Empirical trading results on the Hong Kong stock market (I).

<table>
<thead>
<tr>
<th>Model</th>
<th>Co-integrated LMAR</th>
<th>RADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Financial sector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average days to open</td>
<td>9.94</td>
<td>13.45</td>
</tr>
<tr>
<td>Average days to close</td>
<td>11.57</td>
<td>17.50</td>
</tr>
<tr>
<td>% forced to close</td>
<td>9.09</td>
<td>27.27</td>
</tr>
<tr>
<td>Single trading return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (%)</td>
<td>4.16</td>
<td>4.12</td>
</tr>
<tr>
<td>Median (%)</td>
<td>5.00</td>
<td>6.43</td>
</tr>
<tr>
<td>SD (%)</td>
<td>5.49</td>
<td>5.57</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.80</td>
<td>−0.80</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.77</td>
<td>2.30</td>
</tr>
<tr>
<td>% of observations &lt; 0 (%)</td>
<td>18.18</td>
<td>18.18</td>
</tr>
<tr>
<td>Portfolio return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual return (%)</td>
<td>14.43</td>
<td>5.57</td>
</tr>
<tr>
<td>Annualized volatility (%)</td>
<td>13.81</td>
<td>16.27</td>
</tr>
<tr>
<td>(B) Conglomerate sector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average days to open</td>
<td>8.96</td>
<td>8.58</td>
</tr>
<tr>
<td>Average days to close</td>
<td>7.32</td>
<td>16.75</td>
</tr>
<tr>
<td>% forced to close</td>
<td>18.52</td>
<td>33.33</td>
</tr>
<tr>
<td>Single trading return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (%)</td>
<td>6.05</td>
<td>4.08</td>
</tr>
<tr>
<td>Median (%)</td>
<td>5.79</td>
<td>3.48</td>
</tr>
<tr>
<td>SD (%)</td>
<td>4.74</td>
<td>3.81</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.67</td>
<td>−0.58</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.18</td>
<td>3.91</td>
</tr>
<tr>
<td>% of observations &lt; 0 (%)</td>
<td>11.11</td>
<td>16.67</td>
</tr>
<tr>
<td>Portfolio return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual return (%)</td>
<td>12.93</td>
<td>5.15</td>
</tr>
<tr>
<td>Annualized volatility (%)</td>
<td>24.52</td>
<td>9.39</td>
</tr>
</tbody>
</table>

### Table 4. Empirical trading results on the Hong Kong stock market (II).

<table>
<thead>
<tr>
<th>Model</th>
<th>Co-integrated LMAR</th>
<th>RADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C) Consumer goods sector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average days to open</td>
<td>8.89</td>
<td>6.05</td>
</tr>
<tr>
<td>Average days to close</td>
<td>11.85</td>
<td>29.80</td>
</tr>
<tr>
<td>% forced to close</td>
<td>18.29</td>
<td>31.82</td>
</tr>
<tr>
<td>Single trading return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (%)</td>
<td>6.52</td>
<td>9.68</td>
</tr>
<tr>
<td>Median (%)</td>
<td>7.26</td>
<td>11.59</td>
</tr>
<tr>
<td>SD (%)</td>
<td>15.00</td>
<td>6.32</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.55</td>
<td>−0.79</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.36</td>
<td>2.69</td>
</tr>
<tr>
<td>% of observations &lt; 0 (%)</td>
<td>18.29</td>
<td>13.64</td>
</tr>
<tr>
<td>Portfolio return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual return (%)</td>
<td>18.74</td>
<td>10.74</td>
</tr>
<tr>
<td>Annualized volatility (%)</td>
<td>34.60</td>
<td>14.89</td>
</tr>
<tr>
<td>(D) Stocks in HSI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average days to open</td>
<td>8.41</td>
<td>9.79</td>
</tr>
<tr>
<td>Average days to close</td>
<td>11.86</td>
<td>12.20</td>
</tr>
<tr>
<td>% forced to close</td>
<td>23.21</td>
<td>28.57</td>
</tr>
<tr>
<td>Single trading return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (%)</td>
<td>2.90</td>
<td>3.06</td>
</tr>
<tr>
<td>Median (%)</td>
<td>3.29</td>
<td>3.39</td>
</tr>
<tr>
<td>SD (%)</td>
<td>4.16</td>
<td>7.43</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.85</td>
<td>−1.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.76</td>
<td>5.58</td>
</tr>
<tr>
<td>% of observations &lt; 0 (%)</td>
<td>17.86</td>
<td>14.29</td>
</tr>
<tr>
<td>Portfolio return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual return (%)</td>
<td>36.63</td>
<td>5.94</td>
</tr>
<tr>
<td>Annualized volatility (%)</td>
<td>11.54</td>
<td>6.33</td>
</tr>
</tbody>
</table>
A simulation study shows that the two-step LSE-EM estimating procedure and the BIC are suitable tools for estimating the co-integrated LMAR model. Simulated trading exercises, using the co-integrated LMAR model, in comparison with the regression-ADF-test model, are conducted in several sectors and composite stocks of the Hong Kong Hang Seng Index. The empirical results suggest that the co-integrated LMAR model shortens the converging period of the basket position, which increases capital efficiency, hence generating higher returns by taking advantage of the ‘mean-reverting’ effect, which is accompanied by a reasonably higher volatility in the return.

Acknowledgements

The research of Philip L.H. Yu and W.K. Li is supported by the HKU Small Project Funding (200707176133). We thank the three referees and the Editor for suggestions that led to improvement of the paper. We thank also Ms Vicki Geall and Dr Andrew Carverhill for their help in polishing the paper. W.K. Li also acknowledges HK GRF grant HKU7036/06P for partial support.

References


Dunis, C.L. and Ho, R., Cointegration portfolios of European equities for index tracking and market neutral strategies. J. Asset Mgmt, 2005, 6(1), 33–52.


Appendix A: Proof of the stationarity of the LMAR model

Proof:

The first part of the proof is to embed the time series defined in (1) into a Markov Model. We follow Bec et al. (2008).

Suppose a fixed , , and the indicator of regime 1, , we have

\[ a_t = Z_t \left( \varphi_{1,0} + \sum_{i=1}^{m} \varphi_{1,i} a_{t-i} + \sigma_1 w_t \right) + (1 - Z_t) \left( \varphi_{2,0} + \sum_{i=1}^{m} \varphi_{2,i} a_{t-i} + \sigma_2 w_t \right) \]

where

\[ Z_t = \max(\varphi_{1,0}, \varphi_{2,0}, n) \]

and define \( a_{1-1} = (1, a_{1-1}, \ldots, a_{1-1}) \), \( \xi_1 = (\varphi_{1,0}, \varphi_{1,1}, \ldots, \varphi_{1,m}, 0, \ldots, 0) \) and \( \xi_2 = (\varphi_{2,0}, \varphi_{2,1}, \ldots, \varphi_{2,m}, 0, \ldots, 0) \). We transform the above formula into

\[ a_t = Z_t (\xi_{a-1} + \sigma_1 w_t) + (1 - Z_t) (\xi_{a-1} + \sigma_2 w_t) \] (A1)

Further, let \( p_{1,1} = p(a_{1-1}) \) and \( p_{2,1} = 1 - p(a_{1-1}) \). We have the Markov chain \( a_1 \) defined on \( \mathbb{R}^{1+} \) and, conditional on \( a_{1-1} \), \( a_t \) has a density function \( f(a_t | a_{1-1}) \) given by

\[ f(a_t | a_{1-1}) = p(a_{1-1}) \varphi \left( \frac{a_t - \xi_{a-1}}{\sigma_1} \right) + (1 - p(a_{1-1})) \varphi \left( \frac{a_t - \xi_{a-1}}{\sigma_2} \right) \]
where \( \phi(.) \) is the density function of the standard normal distribution. From this, we have the \( n \)th step transition density in a straightforward way:

\[
h(a_{t+n} | a_t) = \prod_{i=1}^{n} f(a_{t+i} | a_{t+i-1}).
\]

Let \( B^{n+1} \) be the Borel \( \sigma \)-algebra on \( R^{n+1} \); from proposition A1.7 of Tong (1990), the Markov chain \( \{ a_t \} \) is irreducible and aperiodic and the compact sets in \( B^{n+1} \) are small. Therefore, we will apply the drift criterion in theorem 15.0.1 of Meyn and Tweedie (1993) to \( \{ a_t \} \) in the next step to prove ergodicity. The theorem was first obtained by Feigin and Tweedie (1985) based on the idea of a small set (Meyn and Tweedie, 1993, p. 106), and is stated as follows.

**Theorem A.1**: Suppose that \( \{ X_t \} \) is a Markov chain that is irreducible and aperiodic and whose compact sets are ‘small’, and that there exists a measure \( \phi \) and a compact set \( A \) with \( \phi(A) > 0 \) such that

(i) \( \{ X_t \} \) is \( \phi \)-irreducible, and

(ii) there exists a non-negative continuous function \( g \) satisfying

\[
g(x) \geq 1, \quad \text{for } x \in A,
\]

and for some \( \delta > 0,
\]

\[
E[g(X_{t-1}) | X_{t-1} = x] \leq (1 - \delta) g(x), \quad x \in A^c.
\]

Then \( \{ X_t \} \) is geometrically ergodic.

To find the drift function \( g \), we further embed (A1) into

\[
a_t = Z_t \left( \Gamma_1 a_{t-1} + \begin{pmatrix} 0 \\ \sigma_1 w_t \end{pmatrix} \right)
+ (1 - Z_t) \left( \Gamma_2 a_{t-1} + \begin{pmatrix} 0 \\ \sigma_2 w_t \end{pmatrix} \right). \tag{A2}
\]

A popular choice for the drift function that indicates the existence of a second-order moment is proposed by Feigin and Tweedie (1985):

\[
g(x) = 1 + x'Dx \geq 1, \quad D = \sum_{i=0}^{\infty} \Gamma_i^T \Gamma_i.
\]

Because regime 1 is stationary, \( \rho(\Gamma_1) \), which is the largest eigenvalue of \( \Gamma_1 \), is smaller than 1. Therefore, matrix \( D \) is well defined. To verify the conditions, we need to calculate

\[
E(g(a_t) | a_{t-1} = v).
\tag{A3}
\]

However, due to the mixture in the variance part \( (\sigma_1 \neq \sigma_2) \), expression (A3) is very complicated. Note that (A2) is equivalent to

\[
a_t = Z_t \Gamma_1 a_{t-1} + (1 - Z_t) \Gamma_2 a_{t-1} + Z_t \begin{pmatrix} 0 \\ \sigma_1 w_t \end{pmatrix}
+ (1 - Z_t) \begin{pmatrix} 0 \\ \sigma_2 w_t \end{pmatrix}. \tag{A4}
\]

Also note the second-order moment nature of \( E(g(a_t) | a_{t-1} = v) \), and let \( \sigma_0 = \max(\sigma_1, \sigma_2) \). Define a similar Markov chain \( \{ b_t \} \) like (A4) but with volatility in both regimes equal to \( \sigma_0 \) as follows:

\[
b_t = Z_t \Gamma_t b_{t-1} + (1 - Z_t) \Gamma_2 b_{t-1} + \begin{pmatrix} 0 \\ \sigma_0 w_t \end{pmatrix}. \tag{A5}
\]

Clearly, the Markov chain \( \{ b_t \} \) has a larger second moment than \( \{ a_t \} \). Thus, with the same initial value, \( b_{t-1} = a_{t-1} = v \),

\[
E(g(a_t) | a_{t-1} = v) \leq E(g(b_t) | b_{t-1} = v) = 1 + E(b_t'Db_t | b_{t-1} = v).
\]

Let \( f(b_{t-1}) := \Gamma_1 b_{t-1} - (1 - Z_t)(\Gamma_1 - \Gamma_2)b_{t-1} = Z_t \Gamma_t b_{t-1} + (1 - Z_t) \Gamma_2 b_{t-1} \), then we have

\[
E(b_t'Db_t | b_{t-1} = v)
= E \left[ \left( f(v) + \begin{pmatrix} 0 \\ \sigma_0 w_t \end{pmatrix} \right)' D \left( f(v) + \begin{pmatrix} 0 \\ \sigma_0 w_t \end{pmatrix} \right) \right]
= E(f(v)'Df(v)) + E \left[ \begin{pmatrix} 0 \\ \sigma_0 w_t \end{pmatrix}' D \begin{pmatrix} 0 \\ \sigma_0 w_t \end{pmatrix} \right]
= E(f(v)'Df(v)) + \sigma_0^2 d_{22},
\]

where \( d_{22} \) is the element of the second row and second column in matrix \( D \). Further, we have

\[
E(f(v)'Df(v)) = v' \Gamma_t^T D \Gamma_1 v + v' (\Gamma_1 - \Gamma_2)' D(\Gamma_1 - \Gamma_2) v
\times E(1 - Z_t)^2 - 2v' \Gamma_t^T D(\Gamma_1 - \Gamma_2) v E(1 - Z_t).
\]

Since

\[
v' \Gamma_t^T D \Gamma_1 v = \left( \sum_{i=1}^{\infty} \Gamma_t^T \Gamma_i \right) v = v'Dv - v',
\]

and

\[
E(1 - Z_t) = E(1 - Z_t)^2 = 1 - p(v),
\]

we have

\[
E(g(b_t) | b_{t-1} = v) = 1 + \sigma_0^2 d_{22} + (v'Dv - v') + (1 - p(v))d(v)
= g(v) + \sigma_0^2 d_{22} - v' + (1 - p(v))d(v)
\]

\[
= g(v) \left( 1 - \frac{v' \sigma_0^2 d_{22} - (1 - p(v))d(v)}{g(v)} \right),
\]

where

\[
d(v) = v' (\Gamma_1 - \Gamma_2)' D(\Gamma_1 - \Gamma_2) v - 2v' \Gamma_t^T D(\Gamma_1 - \Gamma_2) v.
\]

First, it is clear that

\[
\frac{\sigma_0^2 d_{22}}{g(v)} \to 0, \quad \text{as } g(v) \to \infty.
\]
Also, by definition, \( g(v) = O(\|v\|^2) \) and \( d(v) = O(\|v\|^2) \).

Next, from the conditions in theorem 2.1, \( r_t \to +\infty \) as \( \|v\|^2 \to \infty \), which means \( p(v) \to 1 \) as \( \|v\|^2 \to \infty \). Thus we have

\[
\frac{(1 - p(v))d(v)}{g(v)} \to 0,
\]
as \( \|v\|^2 \to \infty \). Therefore, for \( \lambda > 1 \) we define the compact set

\[ K = \{ v \in \mathbb{R}^{4+1} | v'Dv \leq \lambda \}. \]

Then, for \( \lambda \) large enough, and \( v \in K^c \)

\[
\frac{v'v - \sigma_0^2d_{22} - (1 - p(v))d(v)}{g(v)} \geq \inf \left( \frac{v'v}{g(v)} \right) - \left( \frac{\sigma_0^2d_{22}}{g(v)} + \frac{(1 - p(v))d(v)}{g(v)} \right) 
\]

\[
\geq \inf \left( \frac{v'v}{2\nu'Dv} \right) - \left( \frac{\sigma_0^2d_{22}}{g(v)} + \frac{(1 - p(v))d(v)}{g(v)} \right) 
\]

\[
\geq \frac{1}{2\rho(D)} - (\epsilon_1 + \epsilon_2),
\]

where \( \epsilon_1 \) and \( \epsilon_2 \) are very small numbers. Thus, setting \( 1/(2\rho(D)) - (\epsilon_1 + \epsilon_2) > \delta > 0 \), we have

\[
E(g(a_t | a_{t-1} = v) < E(g(b_t | b_{t-1} = v) \leq (1 - \delta)g(v)).
\]

Thus, we have proved that \( \{a_t\} \), and hence \( a_t \), is geometrically ergodic with finite second-order moment. \( \square \)