Discussion Papers in Economics

No. 05/2016

Pairs trading with partial cointegration

Matthew Clegg
Independent

Christopher Krauss
University of Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
http://www.iwqw.wiso.uni-erlangen.de/forschung/iwqw-discussion-paper-series.shtml

Pairs trading with partial cointegration

Matthew Clegg[a,1], Christopher Krauss[b,1]

[a] *Independent*
[b] *University of Erlangen-Nürnberg, Department of Statistics and Econometrics, Lange Gasse 20, 90403 Nürnberg, Germany*

Wednesday 4[th] May, 2016

Abstract

Partial cointegration is a weakening of cointegration that allows for the "cointegrating" process to contain a random walk and a mean-reverting component. We derive its representation in state space, provide a maximum likelihood based estimation routine, and a suitable likelihood ratio test. Then, we explore the use of partial cointegration as a means for identifying promising pairs and for generating buy and sell signals. Specifically, we benchmark partial cointegration against several classical pairs trading variants from 1990 until 2015, on a survivor bias free data set of the S&P 500 constituents. We find annualized returns of more than 12 percent after transaction costs. These results can only partially be explained by common sources of systematic risk and are well superior to classical distance-based or cointegration-based pairs trading variants on our data set.

Keywords: Statistical arbitrage, pairs trading, quantitative strategies, cointegration, partial cointegration.

Email addresses: matthewcleggphd@gmail.com (Matthew Clegg), christopher.krauss@fau.de (Christopher Krauss)

[1]The authors have benefited from many helpful discussions with Thomas Fischer, Michael Grottke, Ingo Klein, Jerzy Pawlowski, and Johannes Stübinger.

1. Introduction

Pairs trading is a relative-value arbitrage strategy that has been known in the quantitative finance community ever since the mid 1980s (Vidyamurthy, 2004). The strategy involves identifying two securities whose prices tend to travel together. Upon divergence, the cheaper security is bought long and the more expensive one is sold short. When the prices converge back to their historical equilibrium, the trade is closed and a profit collected. By formalizing these concepts, we obtain a quantitative trading strategy exploiting relative mispricings between two securities.

Pairs trading has been introduced to the academic community with the seminal papers of Gatev et al. (1999, 2006). The authors study a pragmatic algorithm on U.S. CRSP securities from 1962 until 2002. At first, they rank all pairs in a 12-month formation period by their sum of squared distances in normalized price space. Then, the top 20 pairs with minimum distance metric are transferred to a subsequent six-month trading period. A trade is entered when the spread diverges at least two historical standard deviations from equilibrium. It is closed with the next zero-crossing, at the end of the trading period, or upon delisting. Gatev et al. (2006) find statistically and economically significant excess returns of 11 percent p.a., which exhibit low exposure to common sources of systematic risk.

Subsequently, interest in pairs trading has surged. Recently, Krauss (2015) has surveyed the literature, cataloging more than 90 papers in five different categories - among them the cointegration approach. Representative studies are by Caldeira and Moura (2013); Huck (2015); Rad et al. (2015). These applications identify cointegrated pairs in a formation period and trade the cointegrating process in a subsequent trading period. Further studies describe the spread in state space, but still as fully mean-reverting model - see, among others, Elliott et al. (2005); Do et al. (2006); Triantafyllopoulos and Montana (2011); de Moura et al. (2016). The results are convincing. For example, Rad et al. (2015), the most comprehensive study, report annualized excess returns of approximately 10 percent from 1962 until 2014 on U.S. CRSP data, prior to transaction costs.

Despite these positive findings, there is reason to believe that cointegration may not be the most appropriate model for pairs trading. If the price series of two securities are cointegrated, then all shocks to the spread series must necessarily be transient. However, it is reasonable to believe that companies often experience idiosyncratic shocks that are permanent in nature. Such shocks might be due to technological changes, the development of new markets, adverse legal actions, changes in management, changes in credit rating, and analyst upgrades or downgrades, to name a few. Any change to the fundamental value of a company arguably represents a permanent shock. This view is supported by three recent contributions: Clegg (2014) finds that the property of being cointegrated is not persistent, and Krauss et al. (2015) as well as Jondeau et al. (2015) identify a substantial number of permanent jumps in high-frequency market data. In light of these findings, it seems fair to assume that the spread series of equity pairs are bombarded by a steady stream of permanent shocks. As of today, no pairs trading study addresses this issue.

Our paper aims to fill this void. We develop the concept of *partial cointegration* (PCI), a weakening of cointegration (CI) that allows for the "cointegrating" process to contain a random walk component.[2] In other words, we model the spread series as the sum of a mean-reverting component and a random walk component. In contrast to Bertram (2010a), we assume that the mean-reverting and random walk components are not directly observable. Even though they are not observable, we are able to show that the system is identifiable, and we give a procedure for estimating the two components. Moreover, if the magnitude of the mean-reverting component is large in comparison to the random walk component, it is possible to profitably trade from the relationship.

Overall, we make three contributions to the literature. The first one is conceptual, i.e., the representation of the PCI model in state space and the discussion of corresponding estimation and testing routines. The second one is simulative. Specifically, we evaluate the goodness of our maximum likelihood based estimators and analyze size as well as power properties of a likelihood ratio test for partial cointegration - originally proposed in Clegg (2015). In a further Monte Carlo simulation, we benchmark PCI-based versus CI-based pairs trading on artificially generated pairs, thereby extracting key determinants of profitability for the

[2] Please note that the term partial cointegration is sometimes used in a different context. Harbo et al. (1998) refer to partial cointegration in case of conditionally cointegrated systems and Caner and Hansen (2001); Krishnakumar and Neto (2009) in the context of partial unit roots in threshold autoregressive models. Both concepts are unrelated to the model presented in this paper.

individual systems. The third contribution is empirical, i.e., the application of cointegration-based pairs trading in comparison to PCI-based pairs trading on the S&P 500 constituents from January 1990 until October 2015.

Our findings are as follows. On simulated data, we identify a key prerequisite for PCI-based pairs trading profitability: The mean-reverting component needs to dominate the random walk component, i.e., the portion of variance attributable to mean reversion has to be relatively high. Then, profitability is a function of the half-life of mean-reversion - the faster the series reverts to its mean, the higher are the returns that can be achieved. On empirical data, we find PCI-based pairs trading to produce annualized returns of more than 12 percent after transaction costs, which are largely robust to common sources of systematic risk. In contrast, classical distance and cointegration-based pairs trading variants disappoint on our highly liquid and survivor bias free data sample.

The remainder of this paper is organized as follows. In section 2, we develop the partial cointegration model and simulatively evaluate the goodness of our estimators and power as well as size properties of the likelihood ratio test. In section 3, we describe the research design for comparing PCI with CI in a pairs trading context - both on simulated and empirical data. The results are presented in section 4. Finally, section 5 provides concluding thoughts and summarizes directions for further research.

2. Partial cointegration

2.1. Representation

The partial cointegration model is a weakening of cointegration that allows for the residual series to have both mean-reverting and random walk components. In analogy with Engle and Granger (1987), the following definition is given:

DEFINITION: The components of the vector X_t are said to be *partially cointegrated of order d, b,* denoted $X_t \sim PCI(d, b)$, if (i) all components of X_t are $I(d)$; (ii) there exists a vector $\alpha (\neq 0)$ so that $Z_t = \alpha' X_t$ and Z_t can be decomposed as a sum $Z_t = R_t + M_t$, where $R_t \sim I(d)$ and $M_t \sim I(d - b)$.

In the remainder of this article, we consider the simplest possible instance of partial cointegration. Namely, we consider only two price time series $X_1 = (X_{1,t})_{t \in T}$ and

$X_2 = (X_{2,t})_{t \in T}$, and we say that X_1 and X_2 are partially cointegrated if $\beta, \rho, \sigma_M, \sigma_R, m_0$, and r_0 can be found such that the following model is satisfied,

$$\begin{aligned}
X_{2,t} &= \beta X_{1,t} + W_t \\
W_t &= M_t + R_t, \\
M_t &= \rho M_{t-1} + \varepsilon_{M,t} \qquad \varepsilon_{M,t} \sim \mathcal{N}\left(0, \sigma_M^2\right), \\
R_t &= R_{t-1} + \varepsilon_{R,t} \qquad \varepsilon_{R,t} \sim \mathcal{N}\left(0, \sigma_R^2\right),
\end{aligned} \tag{1}$$

where $\beta \in \mathbb{R}$ is a parameter, $\rho \in (-1, 1)$ is the AR(1) coefficient, and $\varepsilon_{M,t}, \varepsilon_{R,t}$ follow mutually independent Gaussian white noise processes with expectation zero and variances $\sigma_M^2, \sigma_R^2 \in \mathbb{R}_0^+$. To simplify model estimation, we take $m_0 = 0$ and $r_0 = X_{2,0} - \beta X_{1,0}$. Note that the time series X_2 and X_1 are connected by a partially autoregressive (PAR) model $W = (W_t)_{t \in T}$, first discussed in Summers (1986) and Poterba and Summers (1988), and further elaborated in Clegg (2015) and the associated R package `partialAR`. A key statistic of a PAR model is the proportion of variance attributable to mean-reversion, given as,

$$R_{MR}^2 = \frac{\mathrm{Var}\left[(1-B)M_t\right]}{\mathrm{Var}\left[(1-B)W_t\right]} = \frac{2\sigma_M^2}{2\sigma_M^2 + (1+\rho)\sigma_R^2}, \qquad R_{MR}^2 \in [0,1], \tag{2}$$

where B denotes the backshift operator. If $R_{MR}^2 = 0$, then the AR component is zero and the series a pure random walk. By contrast, if $R_{MR}^2 = 1$, then the random walk component is zero and the series is (fully) autoregressive.

Since W_t is not directly observable, we restate the model in state space. Brockwell and Davis (2010) provide an introductory treatment of state space models, while Durbin and Koopman (2012) offer a more comprehensive reference. The state space representation involves two equations, an *observation equation* and a *state equation*. These equations are customarily given as

$$\boldsymbol{X_t} = \boldsymbol{H_t Z_t} + \boldsymbol{V_t} \tag{3}$$

$$\boldsymbol{Z_t} = \boldsymbol{F_t Z_{t-1}} + \boldsymbol{G_t U_t} + \boldsymbol{W_t}. \tag{4}$$

The state of the system is given by $\boldsymbol{Z_t}$ in (4), which may not be directly observable. It is assumed to follow a linear dynamic and it may be influenced by a control input $\boldsymbol{U_t}$. The term

$\boldsymbol{W_t}$ is a noise term, which has covariance matrix $\boldsymbol{Q_t}$. The observable portion of the system is represented by $\boldsymbol{X_t}$ in (3). It is assumed to have a linear dependence on the hidden state $\boldsymbol{Z_t}$, given by $\boldsymbol{H_t}$, and to be influenced by its own noise term $\boldsymbol{V_t}$, whose covariance matrix is $\boldsymbol{R_t}$. We hereafter assume that the noise term $\boldsymbol{V_t}$ is zero and that there is no control input term $\boldsymbol{U_t}$. In addition, we assume that the linear dependence matrix $\boldsymbol{H_t}$ and the transition matrix $\boldsymbol{F_t}$ are time invariant. Consequently, these equations simplify to

$$\boldsymbol{X_t} = \boldsymbol{H}\boldsymbol{Z_t} \tag{5}$$

$$\boldsymbol{Z_t} = \boldsymbol{F}\boldsymbol{Z_{t-1}} + \boldsymbol{W_t}. \tag{6}$$

The partial cointegration (PCI) system has two observable variables, X_1 and X_2, and two hidden state variables M and R. For convenience of representation, we treat X_1 as a third hidden state variable. In other words, X_1 is represented in both the observation equation and the state equation. The observation equation for the PCI system is therefore given as

$$\boldsymbol{X_t} = \begin{bmatrix} X_{2,t} \\ X_{1,t} \end{bmatrix} = \boldsymbol{H}\boldsymbol{Z_t} = \begin{bmatrix} \beta & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ M_t \\ R_t \end{bmatrix}. \tag{7}$$

And the hidden state equation for the PCI system is given as

$$\boldsymbol{Z_t} = \begin{bmatrix} X_{1,t} \\ M_t \\ R_t \end{bmatrix} = \boldsymbol{F}\boldsymbol{Z_{t-1}} + \boldsymbol{W_t} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ M_{t-1} \\ R_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{X,t} \\ \varepsilon_{M,t} \\ \varepsilon_{R,t} \end{bmatrix}, \tag{8}$$

with $\varepsilon_{X,t} \in \mathbb{R}_0^+$ denoting the variance of the process X_1 in first differences. It is assumed that $\varepsilon_{X,t} \sim \mathcal{N}(0, \sigma_X^2)$ and that it is independent of $\varepsilon_{M,t}$ and $\varepsilon_{R,t}$.

2.2. Estimation of a partial cointegration model

In AppendixA, we give a proof that the PCI system is identifiable. In other words, given a (possibly infinite) realization of a PCI system, there is a unique set of parameter values β, ρ, σ_M and σ_R that give rise to that realization. In this subsection, we turn to the problem of how to estimate these values.

Parameter values are determined through maximum likelihood estimation of the associated Kalman filter. If the parameters of the system are known and the innovations $\varepsilon_{X,t}, \varepsilon_{M,t}$, and $\varepsilon_{R,t}$ are zero-mean Gaussian, uncorrelated, and white, the Kalman Filter minimizes the mean-squared error of the estimated parameters. If the innovations are zero-mean, uncorrelated, and white, but non-Gaussian, then the Kalman Filter is still the best linear estimator (Simon, 2006).

The maximum likelihood method is a standard technique for parameter estimation. Let $\boldsymbol{\Theta_t}$ denote the information that is available up to and including time t, and let $\boldsymbol{\Phi}$ denote the parameter values $\beta, \rho, \sigma_X, \sigma_M$, and σ_R. The one-step ahead prediction error given by the Kalman filter is $\boldsymbol{e_t} = \boldsymbol{X_t} - \boldsymbol{E}[\boldsymbol{X_t}|\boldsymbol{\Theta_{t-1}}, \boldsymbol{\Phi}]$. In case of the PCI model, it can be shown to be

$$\boldsymbol{e_t} = \begin{bmatrix} \beta\varepsilon_{X,t} + \varepsilon_{M,t} + \varepsilon_{R,t} \\ \varepsilon_{X,t} \end{bmatrix}. \tag{9}$$

Since $p(\beta\varepsilon_{X,t} + \varepsilon_{M,t} + \varepsilon_{R,t}, \varepsilon_{X,t}) = p(\varepsilon_{M,t} + \varepsilon_{R,t}, \varepsilon_{X,t})$, the likelihood function for the Kalman filter of the PCI model can be written as

$$\mathcal{L}(\boldsymbol{\Phi}) = p(\boldsymbol{X_1}|\boldsymbol{\Phi}) \prod_{k=2}^{n} \phi(\varepsilon_{M,k} + \varepsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2) \prod_{k=2}^{n} \phi(\varepsilon_{X,k}; 0, \sigma_X^2), \tag{10}$$

where $\phi(\cdot)$ denotes the probability density function of the normal distribution and $p(\boldsymbol{X_1}|\boldsymbol{\Phi})$ is a constant term corresponding to the first observation. We are only interested in optimizing for β, ρ, σ_M, and σ_R, so we can omit the first and third term from the above product. In other words, the maximum likelihood estimates for β, ρ, σ_M and σ_R can be found by maximizing

$$\mathcal{L}_{MR}(\beta, \rho, \sigma_M, \sigma_R) = \prod_{k=2}^{n} \phi(\varepsilon_{M,k} + \varepsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2). \tag{11}$$

We use this likelihood score as the objective function, and deploy L-BFGS to jointly optimize over β, ρ, σ_M, and σ_R. The full algorithm is available online in the R package `partialCI` and the full derivation of the likelihood function is provided in AppendixB.

2.3. Consistency of estimation routine

Given that we employ a maximum likelihood routine, our estimators should be asymptotically efficient, asymptotically unbiased, and asymptotically consistent with regard to mean squared error (MSE), if regularity conditions hold. We check these properties on a collection of synthetic data sets of PCI models with different parameter configurations. The baseline parameter settings are depicted in table 1.

β	σ_X	σ_M	σ_R	ρ	m_0	r_0
1.0000	0.0236	1.0000	1.0000	0.9000	0.0000	0.0000

Table 1: Baseline parameter settings for Monte Carlo simulations.

We set β to one, reflecting a typical equilibrium relationship. The price time series X_t is generated as cumulative return time series with a starting price of 100, mean zero, and standard deviation $\sigma_X = 0.0236$ - corresponding to the median value of all stocks having ever been a S&P 500 constituent from 1990 until 2015.[3] The standard deviations σ_M and σ_R are both set to one for the sake of simplicity. Not their level is important, but their ratio, as it affects the proportion of variance attributable to mean-reversion defined in equation (2). The AR(1)-coefficient ρ is fixed ad hoc at 0.90 in the baseline setting, corresponding to a half-life of mean-reversion of 6.60 days. We now vary *ceteris paribus* either σ_M or ρ, as depicted in table 2, thus reflecting different levels of R^2_{MR} and mean-reversion strength. The sample size n is set to 100, 1000, or 10000 and we perform 10000 replications for each setting.

σ_M	0.00	0.50	1.00	1.50	2.00
ρ	1.00	0.90	0.80	0.70	0.60

Table 2: Different Monte Carlo settings: Ceteris paribus parameter variations.

The results for varying levels of σ_M are depicted in table 3. We see that MSE converges towards zero for all levels of $\sigma_M > 0$, so our estimators seem to be MSE-consistent. The only exception is the Monte Carlo setting with σ_M equal to zero. In this case, the cointegrating process degenerates to a pure random walk, leading to inconsistent estimators. With increasing R^2_{MR}, the MSE exhibits faster convergence for all parameters, up until $R^2_{MR} = 0.51$, i.e., when σ_M equals σ_R. Average estimates for β are very close to the true value for all sample sizes, suggesting that the estimator is unbiased. Conversely, all other estimators suffer from small sample bias and are only asymptotically unbiased (which is a direct consequence of MSE-consistency). Specifically, σ_R and ρ are often underestimated.

[3]Neither the starting price nor the standard deviation σ_X has any influence on the subsequent simulations, but we have decided to choose realistic values motivated from real data.

			Mean squared error			Average estimate			True value
σ_M	R^2_{MR}	Parameter	100	1000	10000	100	1000	10000	
0.00	0.00	β	0.0021	0.0003	0.0001	1.0000	1.0001	0.9998	1.00
0.00	0.00	ρ	0.5541	0.5052	0.5260	0.5069	0.5801	0.5603	0.90
0.00	0.00	σ_M	0.5823	0.5877	0.5542	0.6697	0.6600	0.6057	0.00
0.00	0.00	σ_R	0.4777	0.4802	0.5074	0.4518	0.4643	0.4692	1.00
0.50	0.21	β	0.0026	0.0003	0.0001	1.0001	1.0002	0.9998	1.00
0.50	0.21	ρ	0.5000	0.4148	0.2324	0.5343	0.6159	0.7134	0.90
0.50	0.21	σ_M	0.2295	0.2052	0.0839	0.7767	0.7242	0.5401	0.50
0.50	0.21	σ_R	0.4691	0.3609	0.1166	0.4926	0.6081	0.8857	1.00
1.00	0.51	β	0.0042	0.0005	0.0002	1.0000	1.0002	0.9997	1.00
1.00	0.51	ρ	0.4102	0.1831	0.0075	0.5842	0.7574	0.8921	0.90
1.00	0.51	σ_M	0.2031	0.1213	0.0083	1.0525	1.0285	1.0100	1.00
1.00	0.51	σ_R	0.4673	0.2007	0.0066	0.5885	0.8276	1.0037	1.00
1.50	0.70	β	0.0070	0.0009	0.0004	1.0000	1.0002	0.9996	1.00
1.50	0.70	ρ	0.3169	0.0566	0.0002	0.6420	0.8492	0.8992	0.90
1.50	0.70	σ_M	0.2596	0.0766	0.0031	1.4419	1.5102	1.5202	1.50
1.50	0.70	σ_R	0.5345	0.1451	0.0060	0.6556	0.9083	1.0061	1.00
2.00	0.81	β	0.0107	0.0014	0.0006	1.0000	1.0003	0.9995	1.00
2.00	0.81	ρ	0.2586	0.0172	0.0001	0.6803	0.8812	0.8996	0.90
2.00	0.81	σ_M	0.3345	0.0460	0.0026	1.8785	2.0216	2.0262	2.00
2.00	0.81	σ_R	0.6507	0.1332	0.0068	0.6967	0.9202	1.0058	1.00

Table 3: Mean squared error and average estimate versus true parameters for varying sample size n and different levels of σ_M, thus different levels of R^2_{MR}.

Table 4 reports the results for varying levels of ρ. As before, for ρ equal to one, the PAR model degenerates to a random walk, resulting in inconsistent estimators. For all other settings, we observe MSE-consistent estimation. Small sample bias decreases with ρ.

2.4. Power and size properties of the likelihood ratio test

By construction, a partially autoregressive series normally contains a unit root, so traditional cointegration tests are not appropriate for testing for partial cointegration. Instead, the likelihood ratio test (LRT) is used to assess the null hypothesis $\mathcal{H}_0^R$ that the spread series is a random walk - see Clegg (2015) for further details. When $\mathcal{H}_0^R$ is rejected, we assume that the cointegrating process either follows a partially autoregressive model of order one or autoregressive model of order one. Details about the test statistic and associated critical

			Mean squared error			Average estimate			True value
ρ	R^2_{MR}	Parameter	100	1000	10000	100	1000	10000	
1.00	0.50	β	0.0041	0.0005	0.0002	1.0003	1.0002	0.9997	1.00
1.00	0.50	ρ	0.6348	0.5629	0.6007	0.5074	0.5832	0.5539	1.00
1.00	0.50	σ_M	0.2760	0.3098	0.4090	0.9374	0.9200	0.8251	1.00
1.00	0.50	σ_R	0.4839	0.4926	0.5447	0.6445	0.6749	0.6978	1.00
0.90	0.51	β	0.0042	0.0005	0.0002	1.0000	1.0002	0.9997	1.00
0.90	0.51	ρ	0.4102	0.1831	0.0075	0.5842	0.7574	0.8921	0.90
0.90	0.51	σ_M	0.2031	0.1213	0.0083	1.0525	1.0285	1.0100	1.00
0.90	0.51	σ_R	0.4673	0.2007	0.0066	0.5885	0.8276	1.0037	1.00
0.80	0.53	β	0.0044	0.0006	0.0002	1.0000	1.0002	0.9997	1.00
0.80	0.53	ρ	0.3295	0.0701	0.0010	0.5626	0.7381	0.7979	0.80
0.80	0.53	σ_M	0.1810	0.0533	0.0030	1.0656	1.0270	1.0223	1.00
0.80	0.53	σ_R	0.4069	0.0693	0.0029	0.6372	0.9518	1.0190	1.00
0.70	0.54	β	0.0045	0.0006	0.0002	1.0000	1.0002	0.9997	1.00
0.70	0.54	ρ	0.2701	0.0270	0.0011	0.5297	0.6742	0.6983	0.70
0.70	0.54	σ_M	0.1622	0.0262	0.0028	1.0767	1.0383	1.0338	1.00
0.70	0.54	σ_R	0.3561	0.0289	0.0027	0.6840	0.9992	1.0325	1.00
0.60	0.56	β	0.0046	0.0006	0.0002	1.0000	1.0001	0.9997	1.00
0.60	0.56	ρ	0.2318	0.0172	0.0012	0.4813	0.5865	0.5985	0.60
0.60	0.56	σ_M	0.1472	0.0178	0.0034	1.0845	1.0498	1.0464	1.00
0.60	0.56	σ_R	0.3078	0.0174	0.0034	0.7335	1.0245	1.0461	1.00

Table 4: Mean squared error and average estimate versus true parameters for varying sample size n and different levels of ρ, thus different levels of R^2_{MR} and half-life of mean-reversion.

values can be found in AppendixC. We follow similar Monte Carlo settings as described in subsection 2.3, but we choose a finer step size for σ_M and ρ.

The results for varying levels of σ_M are depicted in table 5. When σ_M equals zero, i.e., the cointegrating process is a pure random walk, we find that the null is erroneously rejected in approximately four to five percent of all cases. As such, the type I error is in line with the significance level α - size distortions are minor. With increasing levels of σ_M at fixed σ_R, the proportion of variance attributable to mean-reversion increases and consequently, the power of the likelihood ratio test. When $\sigma_M = \sigma_R = 1$, the power is approximately 50 percent for $n = 1000$ - a reasonable value, which could easily be achieved with four years of daily data.

Table 6 reports the results for varying levels of ρ. The random walk case for ρ equal to

σ_M	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
$n = 100$	0.0345	0.0388	0.0490	0.0653	0.0841	0.1033	0.1231	0.1447	0.1666
$n = 1000$	0.0487	0.0569	0.1129	0.2616	0.5064	0.7469	0.8953	0.9625	0.9869
$n = 10000$	0.0428	0.0829	0.5402	0.9750	0.9986	1.0000	1.0000	1.0000	1.0000
R^2_{MR}	0.0000	0.0617	0.2083	0.3719	0.5128	0.6219	0.7031	0.7632	0.8081

Table 5: Power of likelihood ratio test with null hypothesis of "random walk" for varying sample size n and different levels of σ_M, thus different levels of R^2_{MR}.

one results in almost the same size values as before. Power quickly increases with decreasing levels of ρ. For example, for $n = 1000$ and $\rho = 0.80$, the power is already at 87 percent.

ρ	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
$n = 100$	0.0366	0.0598	0.0841	0.1125	0.1402	0.1726	0.2081	0.2472	0.2902
$n = 1000$	0.0452	0.2651	0.5064	0.7254	0.8748	0.9532	0.9848	0.9963	0.9995
$n = 10000$	0.0385	0.9336	0.9986	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R^2_{MR}	-	0.5063	0.5128	0.5195	0.5263	0.5333	0.5405	0.5479	0.5556

Table 6: Power of likelihood ratio test with null hypothesis of "random walk" for varying sample size n and different levels of ρ, thus different levels of R^2_{MR} and half-life of mean-reversion.

3. Study design: Comparing partial cointegration with cointegration in the context of pairs trading

3.1. Data

The backtesting of all trading strategies is performed on the S&P 500 index constituents. As in Krauss et al. (2016), our choice is motivated by market efficiency, computational feasibility, and liquidity. The S&P 500 contains the leading 500 companies of the U.S. stock market, comprising approximately 80 percent of available market capitalization (S&P Dow Jones Indices, 2015). This highly liquid subset serves as an acid test for any trading strategy, in light of significant investor scrutiny and intense analyst coverage. We proceed along the lines of Krauss et al. (2016) for eliminating survivor bias. In particular, we download all month end constituent lists for the S&P 500 from Thomson Reuters Datastream from December 1989 to September 2015. We consolidate these lists into a binary matrix, indicating if the stock is an index constituent or not. Second, for all stocks having ever been in the S&P

500 during this time, we download the daily total return indices (RI) from January 1990 until October 2015. Return indices reflect cum-dividend prices and account for all further corporate actions and stock splits, making it the most adequate metric for return computations. Previously reported concerns about Datastream quality by Ince and Porter (2006) mainly address small size deciles. Also, as Leippold and Lohre (2012) point out, Thomson Reuters seems to have reacted in the meantime. Hence, besides eliminating holidays, we conduct no further sanitizations. In table 3.1, we provide descriptive statistics for our sample period, split into 10 sectors, as defined by the Global Industry Classification Standard (GICS).

Sector	No. of stocks	Arithmetic return	Geometric return	Standard deviation	Skewness	Kurtosis
Industrials	80.6	0.99	0.85	5.36	-0.19	1.71
Consumer Services	72.6	1.07	0.93	5.27	-0.20	2.59
Basic Materials	35.2	0.90	0.70	6.31	-0.02	2.24
Telecommunications	10.7	0.92	0.71	6.50	0.34	4.76
Health Care	41.3	1.33	1.23	4.40	-0.40	1.18
Technology	50.3	1.41	1.05	8.50	-0.06	1.11
Financials	78.0	1.13	0.94	6.17	-0.39	2.44
Consumer Goods	65.2	1.04	0.93	4.53	-0.44	3.02
Oil & Gas	31.2	1.00	0.76	6.89	-0.03	1.06
Utilities	34.6	0.85	0.74	4.54	-0.43	1.72
All	499.7	1.04	0.93	4.78	-0.49	2.01

Table 7: Summary statistics for S&P 500 constituents from January 1990 until October 2015. Returns and standard deviations are denoted in percent.

3.2. The backtesting framework

3.2.1. Building blocks

We construct a backtesting framework to compare CI-based to PCI-based trading. This subsection outlines the building blocks of the backtesting framework. A concise summary of key parameters is provided in table 8.

Strategy variants: In particular, we benchmark four strategy variants against each other. First, we replicate classical distance-based pairs trading of Gatev et al. (2006), denoted as GGR. Second, we run a cointegration-based strategy subsuming proven parameter settings in the available literature (CI1). Third, we test a cointegration-based variant with exactly

	GGR	CI1	CI2	PCI
Number of study periods	305	305	305	305
Formation period, in months	12	12	48	48
Trading period, in months	6	6	6	6
Restriction to same sector	Yes	Yes	Yes	Yes
Eligibility criterion 1	-	ADF test	PP test	LRT in bottom 5%
Eligibility criterion 2	-	Johansen test	PGFF test	$\rho > 0.5$
Eligibility criterion 3	-			$R^2_{MR} > 0.5$
Portfolio size, in pairs	20	20	20	20
Portfolio selection	SSD	In-sample Sharpe ratio	In-sample Sharpe ratio	In-sample Sharpe ratio
Opening signal τ_o	2.0	2.0	1.0	1.0
Closing signal τ_c	0.0	0.0	-0.5	-0.5
Stop loss, in percent	-	-	10	10
Transaction costs, in percent	0.05	0.05	0.05	0.05

Table 8: Summary of backtesting framework.

the same parameter settings as the PCI-system (CI2). Fourth, we run a strategy based on partial cointegration (PCI).

Study periods: For all models, we follow Gatev et al. (2006) and split our data into a collection of overlapping study periods. Each study period consists of a formation period, during which the different models are calibrated, followed by a trading period, during which simulated trading is performed. For the GGR approach, we opt for a 12-month formation, and a 6-month trading period (Gatev et al., 2006). For CI1, we follow Huck (2015) as well as Rad et al. (2015) and use the same durations as GGR. In case of the CI2 and the PCI approach, we lengthen the formation period to 48 months (approximately 1000 days) in order to increase the power of the tests for cointegration and partial cointegration - see tables 5 and 6 as well as Clegg (2014); Krauss et al. (2015). The trading period remains at 6 months. For all four trading systems, the first trading period starts January 1990 and the last trading period ends October 2015. The trading periods are overlapping, so in general, we have six active trading periods in parallel, with the exception of the first five trading periods.

Return computation: Return computation follows Gatev et al. (2006). Specifically, we scale the sum of monthly payoffs across all pairs by the sum of invested capital at the previous month's end. Invested capital is either defined as committed capital (1 USD is allocated at the beginning of the trading period for each pair - whether it opens or not) or fully invested capital (1 USD is allocated at the beginning of the trading period for each pair that opens during the trading period). The former metric yields return on committed capital and the latter metric return on fully invested capital.

3.2.2. Formation period

Available stocks: Within each formation period, a universe of available stocks is determined. For pair formation, we only consider the stocks that (i) are an index constituent of the S&P 500 at the very last day of the formation period and (ii) do not exhibit missing data during the (in-sample) formation period.

Eligible pairs: From the universe of available stocks, a collection of eligible pairs is then determined. Following previous studies (Gatev et al., 2006; Do and Faff, 2010, 2012) we only allow pairs composed of stocks belonging to the same GICS-sector. This restriction renders our application computationally feasible[4] and reduces spurious relationships. For the GGR approach, all pairs of the same sector are eligible and no further restrictions are imposed. For CI1, a same-sector pair is eligible, if both the Augmented Dickey-Fuller test of Dickey and Fuller (1979) and the Johansen test of Johansen (1988) reject the null hypothesis of "no cointegration relationship" at the level $\alpha = 0.05$. These tests are recommended in Huck (2015) and Rad et al. (2015). Caldeira and Moura (2013) use this exact combination. In case of CI2, a same-sector pair is eligible if both the Phillips-Perron test of Phillips and Perron (1988) and the Pantula, Gonzalez-Farias, and Fuller test of Pantula et al. (1994) reject the null hypothesis of "no cointegration relationship" at the level $\alpha = 0.05$. These tests exhibit higher power in case of financial return data - see Clegg (2014) and Krauss et al. (2015) for a comparison of various unit root tests. For PCI, a same-sector pair is eligible if the likelihood

[4]Note: Parallelized processing on 8 hyper threads on a contemporary Intel core i7-4790K with a clock speed of 4 GHz leads to an approximate run-time of 15 days for the PCI-application and 11 hours for each of the other variants, considering industry restrictions.

ratio score is in the lower five percent among all available pairs, $\rho > 0.5$, and $R^2_{MR} > 0.5$. The first condition (likelihood ratio score in the lower 5 percent) ensures that the pairs have a higher probability of being (partially) cointegrated. The second condition ($\rho > 0.5$) excludes pairs with a half-life of mean-reversion of one day or less - thereby avoiding to select pairs where trading gains are largely attributable to bid-ask bounce. The third condition ($R^2_{MR} > 0.5$) ensures more reliable parameter estimates - see tables 3 and 4.

Top pairs: For GGR, the top pairs transferred to the trading period are the ones with minimum sum of squared distance (SSD) in normalized price space - see Gatev et al. (2006) for further details. For CI1, CI2, and PCI we run an in-sample trading simulation adhering to the standards outlined in table 8 and detailed in subsection 3.2.3. Following Dunis et al. (2010); Bertram (2010b); Caldeira and Moura (2013), we assign an in-sample Sharpe ratio to each pair. For each stock, its eligible partners are then ranked by their respective in-sample Sharpe ratios. The highest ranking partner is then selected as the trading partner for that stock. Thus, we obtain a list of securities and their selected trading partners. Finally, we construct a portfolio consisting of the 20 stocks and their partners with the highest in-sample Sharpe ratios - the top pairs selected for the subsequent trading period.

3.2.3. Trading period

In the trading period, we consider the top 20 pairs selected in the formation period. In the following, we discuss the buy and sell signals for the different strategy variants.

Distance-based pairs trading (GGR): For a given pair (P, Q), let X_1 and X_2 represent their respective price series. Using in-sample data of the formation period, we obtain the historical standard deviation σ of the price spread $X_2 - X_1$. In the trading period, we compute the normalized Z-score Z_t as

$$Z_t = \frac{X_{2,t} - X_{1,t}}{\sigma}.$$

If Z_t is larger than the opening threshold τ_o and no position is currently open, we go long in stock P and short in Q. If a long position in P is already open and Z_t becomes smaller than the closing threshold τ_c, the positions are closed. Analogously, if Z_t is smaller than the opening threshold $-\tau_o$ and no position is currently open, we go short in stock P and long in Q. If a long position in Q is already open and Z_t becomes larger than the closing threshold $-\tau_c$, the positions are closed. We choose $\tau_o = 2.0$ and $\tau_c = 0$, and invest an equal-dollar amount in the long and in the short leg of each pair - just as in the original application of Gatev et al. (2006).

Cointegration-based pairs trading (CI1, CI2): Relying on the same notation as before, we use in-sample data of the formation period to obtain in-sample regression constants α and β, that give rise to the least squares fit to the equation,

$$X_{2,t} = \alpha + \beta X_{1,t} + W_t,$$

where W_t denotes the zero-mean residual series. The historical standard deviation σ of W_t is then determined - equally with formation data. In the trading period, we compute the normalized Z-score Z_t as

$$Z_t = \frac{X_{2,t} - \alpha - \beta X_{1,t}}{\sigma},$$

which we use as trading signal. For CI1, we follow Huck (2015); Rad et al. (2015) and adopt an opening threshold $\tau_o = 2$ and a closing threshold $\tau_c = 0$. For CI2, we take on an opening threshold $\tau_o = 1$ and a closing threshold $\tau_c = -0.5$ - identical to the parametrization of the PCI model described in the next paragraph. In case of CI1, we use no stop-loss rule. By contrast, for CI2, we implement a stop loss of 10 percent, following Nath (2003); Caldeira and Moura (2013) and the development of pairs trading profitability over time presented in Jacobs and Weber (2015). Specifically, if the value of the portfolio associated to a pair drops below 10 percent of its initial value, the trade is closed and the pair is disqualified from any further trades in this trading period. Any positions open on the last day of the trading period are closed and stocks that are delisted booked out with the delisting return[5].

The value of the position on the long side may not exactly match the value of the position on the short side. The assumption underlying the cointegration model is that the

[5] By contrast, if a stock drops out of the S&P 500 during the trading period, we continue trading it until the end of the trading period.

spread series W_t is mean-reverting. Therefore, the goal is to construct a financial position whose behavior mimics the behavior of W_t as closely as possible. According to the above cointegrating equation, each share of Q that is bought long (respectively, sold short) should be offset with a short sale (respectively, long purchase) of β shares of P. Thus, a USD 1 investment in Q, representing $1/X_{2,t}$ shares in Q, is offset by an investment of $\beta/X_{2,t}$ shares in P, with a price of $\beta X_{1,t}/X_{2,t}$. In general, this amount is not exactly equal to USD 1.

The value of the portfolio associated to a pair may fluctuate on a daily basis. When no trade is open, the portfolio is assumed to be held in cash and not to earn interest. When a trade is open, the value of the portfolio changes according to changes in the value of two securities. When a trade is initiated, it is assumed that the maximum amount of capital that is allocated to either side of the trade is limited to the value of the portfolio. Thus, if $\beta X_{1,t}/X_{2,t} < 1$ and the value of the portfolio on day t is v, then the amount invested in Q (long or short) is v, while the amount invested in P is $(\beta X_{1,t}/X_{2,t})v$. If $\beta X_{1,t}/X_{2,t} > 1$, then the amount invested in Q is $(X_{2,t}/\beta X_{1,t})v$ while the amount invested in P is v. These restrictions mimic simplified *Reg T* requirements. The total return on the portfolio associated to the pair (P, Q) is the product of the returns achieved for each of the trades in this pair over the trading period.

PCI-based pairs trading (PCI): For each pair (P, Q), the partial cointegrating relationship in the formation period is found:

$$\begin{aligned} X_{2,t} &= \beta X_{1,t} + W_t, \\ W_t &= M_t + R_t, \\ M_t &= \rho M_{t-1} + \varepsilon_{M,t}, \\ R_t &= R_{t-1} + \varepsilon_{R,t}. \end{aligned} \tag{12}$$

In this representation, M_t is the mean-reverting component and hence represents the opportunity for potential trading profit. However, M_t and R_t are not directly visible and need to be estimated using the Kalman filter. Based upon the fitted values for ρ, σ_M, and σ_R, the Kalman gain κ is computed. The Kalman gain can be computed either as a closed-form formula (see Clegg (2015)) or it can be approximated through the Kalman filter equations.

The estimated values of M_t and R_t are then determined through the equations

$$\begin{aligned} W_t &= P_t - \beta Q_t, \\ E_t &= W_t - \rho M_{t-1} - R_{t-1}, \\ M_t &= \rho M_{t-1} + \kappa E_t, \\ R_t &= R_{t-1} + (1-\kappa) E_t, \end{aligned} \tag{13}$$

with E_t representing the one-step estimation error of the Kalman filter. These equations are applied to the data in the formation period to obtain an in-sample estimate of M_t and its standard deviation σ. In the trading period, these equations are again used to estimate M_t at the end of each trading day. A Z-score Z_t is computed as $Z_t = M_t/\sigma$. We take on an opening threshold $\tau_o = 1$ and a closing threshold $\tau_c = -0.5$. These thresholds are close to optimal for PCI on simulated data - see subsections 3.3 and 4.1.

The remainder of the PCI-based trading system mirrors the cointegration-based trading system. In particular, a hedge ratio is determined using the value $\beta X_{1,t}/X_{2,t}$. One remaining significant difference is that under the PCI-based system, it is possible to lose money from convergent trades. In the cointegration-based system, if a trade closes prior to the end of the trading period, it is guaranteed to yield a profit. No such guarantee exists in the partially cointegrated system, due to the random walk component.

3.3. Trading on simulated data

Prior to the empirical application, we aim to explore the effectiveness of trading with the partial cointegration model under ideal circumstances. In particular, the following questions are of interest: First, when the data is partially cointegrated, how does trading using the PCI model compare to trading using the cointegration model? Second, how does the effectiveness of trading vary as a function of the proportion of variance attributable to mean-reversion R^2_{MR} and as a function of ρ, the AR(1)-coefficient? Third, how should optimal entry and exit thresholds τ_o, τ_c be chosen for the PCI system?

To address the first two questions, a collection of synthetic data sets are generated for various values of ρ and R^2_{MR}. The value of ρ is allowed to range from 0.6 to 0.95, while the value for R^2_{MR} is allowed to range from 0.05 to 1.0. For each combination of ρ and R^2_{MR}, 5000 random pairs are generated. To construct such a pair, we follow the procedure outlined in section 2.3 to generate synthetic price time series $X_{1,t}$ of length 1125. Next, a synthetic partially autoregressive process W_t of the same length is generated, subject to the constraint that the variance of the first differences is held at 1.0. Then, the price series $X_{2,t}$ is defined as $X_{2,t} = X_{1,t} + W_t$. Thus, the value of β is fixed at one, assuming an equilibrium relationship between the two securities. Given this pair, trading is then simulated on it using both the cointegrated and the partially cointegrated trading system. The first 1000 observations are used as training data, and trading is then performed on the final 125 observations. This setting roughly corresponds to a four-year formation and a six-month trading period. We benchmark the CI2 strategy against the PCI strategy. Parameters are set as in table 8.

To address the third question, we simulate 5000 random pairs, following the procedure above. However, we choose a parametrization we expect to be typical for an eligible PCI pair, i.e., ρ at 0.9 and R^2_{MR} at 0.75, i.e., a half-life of mean-reversion of 6.6 days and a strong mean-reverting overlay. Next, we run simulated trading on this dataset with $\tau_o \in [0.5, 1.0, \ldots, 2.5]$ and $\tau_c \in [-2.5, -2.45, \ldots, 2.5]$ and log the mean return per month for each configuration of opening and closing threshold. All other parameters are chosen as in table 8.

4. Results

4.1. Simulated data

First, we discuss the trading results in terms of monthly mean return on simulated data. Figure 1 reports the findings. The graph consists of five panels, each representing a different value for ρ. Within each panel, the average monthly return of each strategy is plotted as a function of R^2_{MR}. There are several interesting conclusions we can infer from these graphs.

Neither strategy provides significant gains when $R^2_{MR} < 0.5$, due to the predominant random walk component. With increasing proportion of variance attributable to mean-reversion, we observe increasing returns for both strategies and across all regimes of ρ. This effect is driven by an increasingly strong mean-reverting overlay on top of the random walk component. Higher levels of ρ lead to lower returns for CI-based and PCI-based pairs trading. With higher ρ, the half-life of mean-reversion increases, and as such the trade duration - with a detrimental effect on monthly returns.

Interestingly enough, the level of ρ shifts the relative advantage of CI-based versus PCI-based pairs trading for $R^2_{MR} > 0.5$. For ρ smaller than 0.9, the PCI system clearly outperforms classical cointegration-based pairs trading. This edge disappears for increasingly higher levels of ρ, given that the AR(1)-process then approaches a random walk, so that reliable parameter estimates are harder to achieve. The relative advantage of PCI also disappears for very high levels of R^2_{MR}. For $R^2_{MR} \to 1$, partial cointegration converges towards classical cointegration, rendering the two strategies more similar again.

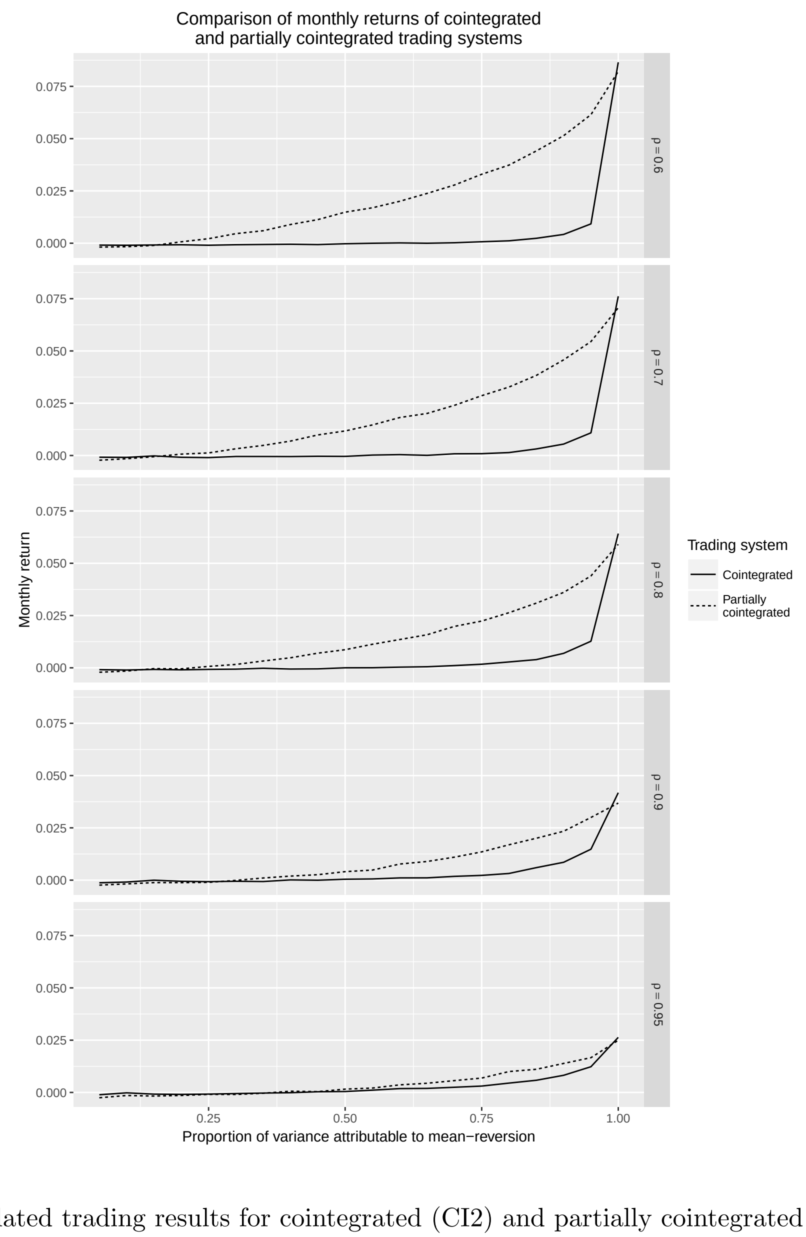

Figure 1: Simulated trading results for cointegrated (CI2) and partially cointegrated strategies (PCI).

Please note that increasing levels of the variance of the first differences of the spread process W_t only influences absolute returns, not the relative advantage of the strategies. To put it simply, it is the same game on a different level. Evidently, σ_X, has no effect either, as it does not have any impact on W_t - see section 2.

For trading applications, we draw three major conclusions: First, only pairs with a high R^2_{MR} should be considered for trading. Second, a PCI-based trading system performs best for intermediate levels of ρ and converges towards the performance of a CI-based system for high ρ and/or high R^2_{MR}. Third, even if the data is only partially cointegrated, it is still possible to profitably trade with a CI-based model.

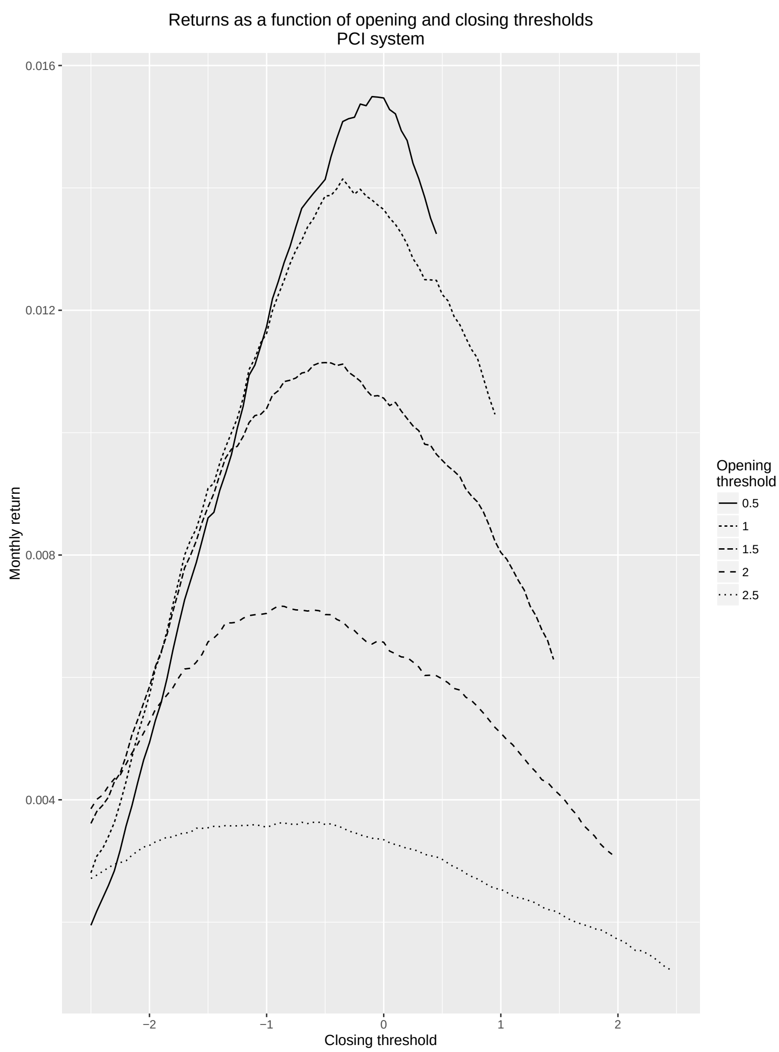

Figure 2: Simulated trading results for partially cointegrated strategy (PCI).

Second, we address the question about optimal trading thresholds for PCI. Figure 2 displays the findings. The chart represents the monthly mean return averaged over 5000 simulated pairs with $\rho = 0.9$ and $R^2_{MR} = 0.75$, evaluated for varying levels of the closing threshold, conditional to different opening thresholds. We observe that monthly mean return decreases with increasing opening threshold. The higher the opening threshold, the higher the average profit per trade, but the lower the number of trades. The relative advantage between the opening thresholds $\tau_o = 0.5$ and $\tau_o = 1.0$ is much smaller than between the opening thresholds $\tau_o = 1.0$ and $\tau_o = 1.5$, due to increasing impact of transaction costs. For practical applications, $\tau_o = 1.0$ seems to be a sensible value, representing a fair trade-off between trading frequency and profit per trade. The simulatively evaluated return maximum for this opening threshold is close to $\tau_c = -0.5$, which we choose for our empirical application.

4.2. Empirical data

We follow Krauss et al. (2016) and run a fully-fledged performance evaluation on all four trading strategies, relying on the R package `PerformanceAnalytics` by Peterson and Carl (2014). Most of the metrics discussed below can be found in Bacon (2008). Our key results are reported in two panels - before and after transaction costs. Given that the strategies are implemented on a highly liquid stock universe, we follow Avellaneda and Lee (2010) and assume transaction costs of 0.05 percent per share per half-turn. This value is well in line with several studies in pairs trading research. For example, Bogomolov (2013) points out that even retail commissions are at 0.10 percent, and Do and Faff (2012) refer to Jones (2002), assuming institutional commissions of 0.10 percent or less between 1997 and 2009. Pairs trading is a liquidity providing strategy exploiting transitory pricing errors, and as such, potentially eligible for rebates (Brogaard et al., 2014). As such, we do not account for any further market impact. In any case, for the S&P 500, market impact would be low. Exact values are hard to estimate, but Prager et al. (2012) report that the bid-ask spread has declined to less that one cent for the S&P 500 constituents, or 2 basis points assuming an average stock price of 50 USD. This value seems reasonable, given that Krauss et al. (2015) find average bid-ask spreads of 4 to 5 basis points on a high frequency data set of the German DAX 30 constituents - also a highly liquid stock universe.

4.2.1. Performance evaluation

Table 9 reports monthly return characteristics and associated risk metrics. We see that the baseline GGR strategy returns 0.15 percent per month prior to transaction costs. This value is reasonable and compares well to the findings of Do and Faff (2010) in the recent part of their CRSP data sample. They achieve 0.56 percent from 1989 until 2002 and 0.33 percent from 2003 until 2009. We surmise that this additional decline is mainly caused by our smaller stock universe, consisting of no more than 500 securities at any point in time (versus several thousands in the case of CRSP data), the extended sample period until 2015 (versus 2002 for Gatev et al. (2006) and 2009 for Do and Faff (2010)), and the higher liquidity of the S&P 500 constituents. Compared to the results of Gatev et al. (2006) of 1.44 percent per month from 1962 until 2002, we can confirm the declining profitability of distance-based pairs trading discussed in Do and Faff (2010). In the spirit of Rad et al. (2015), classical cointegration-based pairs trading performs very similar to distance-based pairs trading with mean returns of 0.20 percent (CI1) and 0.15 percent (CI2) prior to transaction costs. The different cointegration tests and trading thresholds reflected in CI1 and CI2 hardly make any difference. We have not been able to reproduce the higher returns of Huck (2015); Huck and Afawubo (2015) - presumably due to the different data sources (Thomson Reuters versus Bloomberg), the longer sample period (26 versus 10 years), the different treatment of survivor bias, and different pre-selection methodologies. Compared to the classical pairs trading variants, PCI shows clear outperformance with a mean return of 1.18 percent per month before transaction costs - which is large in an economical and a statistical sense (t-statistic = 5.67). This picture barely changes with the inclusion of transaction costs. Even then, the PCI strategy returns 1 percent per month and outperforms a buy and hold investment in the general market (MKT), yielding 0.86 percent per month. Specifically, compared to the general market, the PCI return distribution is skewed to the right - a positive property for any investor and atypical for financial data (Cont, 2001). Also, value at risk (VaR) levels after transaction costs are substantially lower: The historical monthly VaR at the one percent level amounts to -3.5 percent for PCI versus -10 percent for the general market. The reduced tail risk is also expressed in a maximum drawdown of merely 19 percent for PCI, compared to 50 percent for MKT. The hit rates complement this picture

- the share of observations with returns greater than zero is higher than 70 percent for PCI, compared to 64 percent for MKT, and values well below 60 percent for the remaining strategies. We summarize that PCI-based pairs trading exhibits favorable risk and return characteristics - even after transaction cost estimates. However, it remains to be checked whether these returns are robust to systematic sources of risk.

	A: Before transaction costs				B: After transaction costs				
	GGR	CI1	CI2	PCI	GGR	CI1	CI2	PCI	MKT
Mean return (fully invested)	0.0015	0.0020	0.0015	0.0118	0.0010	0.0015	0.0010	0.0100	0.0086
Mean return (committed)	0.0014	0.0018	0.0015	0.0117	0.0009	0.0013	0.0010	0.0100	0.0086
Standard error (NW)	0.0006	0.0009	0.0007	0.0021	0.0006	0.0009	0.0007	0.0020	0.0026
t-Statistic (NW)	2.6149	2.1436	2.0124	5.6729	1.7053	1.5889	1.3762	5.0657	3.2693
Minimum	-0.0522	-0.0531	-0.0791	-0.0699	-0.0528	-0.0535	-0.0794	-0.0687	-0.1708
Quartile 1	-0.0034	-0.0069	-0.0064	-0.0005	-0.0039	-0.0074	-0.0067	-0.0016	-0.0175
Median	0.0015	0.0013	0.0012	0.0092	0.0009	0.0008	0.0004	0.0077	0.0140
Quartile 3	0.0071	0.0102	0.0090	0.0211	0.0066	0.0097	0.0085	0.0188	0.0378
Maximum	0.0295	0.0605	0.0610	0.1960	0.0284	0.0600	0.0595	0.1915	0.1136
Standard deviation	0.0095	0.0161	0.0142	0.0244	0.0095	0.0161	0.0141	0.0238	0.0433
Skewness	-0.8030	0.2435	-0.3102	1.8211	-0.8307	0.2428	-0.3451	1.8718	-0.6498
Kurtosis	3.7743	1.6566	4.5188	11.4527	3.8391	1.6588	4.5286	11.9093	1.1373
Historical VaR 1%	-0.0260	-0.0368	-0.0288	-0.0363	-0.0267	-0.0373	-0.0291	-0.0347	-0.1019
Historical CVaR 1%	-0.0347	-0.0458	-0.0489	-0.0502	-0.0353	-0.0462	-0.0493	-0.0504	-0.1329
Historical VaR 5%	-0.0137	-0.0221	-0.0186	-0.0225	-0.0142	-0.0227	-0.0194	-0.0239	-0.0700
Historical CVaR 5%	-0.0229	-0.0330	-0.0290	-0.0336	-0.0234	-0.0335	-0.0294	-0.0340	-0.0971
Maximum drawdown	0.1057	0.2377	0.2112	0.1297	0.1413	0.2754	0.2544	0.1883	0.5029
Calmar ratio	0.1646	0.0940	0.0798	1.1305	0.0783	0.0586	0.0427	0.6553	0.1910
Share with return > 0	0.5968	0.5484	0.5548	0.7323	0.5677	0.5194	0.5258	0.7032	0.6355

Table 9: Monthly return characteristics and risk metrics of GGR, CI1, CI2, PCI compared to MKT from January 1990 until October 2015. NW denotes Newey-West. The t-statistics are calculated using NW standard errors with six-lag correction.

Table 10 contains summary statistics on trading frequency. Across all systems, almost all pairs are traded at some point during the six-month trading period. The share is slightly lower for GGR and CI1 with opening thresholds of 2.0 (approximately 95 percent), and slightly higher for CI2, and PCI with opening thresholds of 1.0 (approximately 99 percent). Specifically, for the GGR strategy, 94.5 percent of the 20 pairs are traded - a value that compares well with the findings of Gatev et al. (2006) with 19.30 out of 20 pairs, or 96.5

percent. This high value also explains why return on fully invested and on committed capital in table 9 exhibit very similar values. The average number of trades per pair is vastly different for GGR, CI1, CI2 (approximately 1.5), compared to PCI (approximately 6.5). We can cautiously infer from this statistic, that the latter strategy successfully captures the mean-reverting property of the hidden PAR process W_t. In the same vein, almost all PCI-triggered trades are completed during the trading period, i.e., 89 percent, versus levels of 40 to 50 percent for classical pairs trading. Clearly, trade duration of 16.7 days is much lower for PCI-based pairs, reflecting a shorter half-life of mean-reversion.

	GGR	CI1	CI2	PCI
Proportion of pairs traded	0.9446	0.9566	0.9856	0.9860
Average number of trades per pair	1.4108	1.6261	1.6562	6.4800
Average completed trades per pair	0.6444	0.8184	0.7264	5.7500
Share of completed trades per pair	0.4568	0.5033	0.4386	0.8873
Average holding time per trade	60.5594	64.5820	48.2093	16.7000

Table 10: Trading statistics for GGR, CI1, CI2, PCI from January 1990 until October 2015, per six-month trading period.

Table 11 summarizes annualized risk and return characteristics for all four strategies. Annualized return is almost at 15 percent for PCI, prior to transaction costs and at 12.3 percent post transaction costs - well superior to the classical pairs trading variants. Compared to the general market, we observe that we achieve higher returns at approximately half the standard deviation, leading to a favorable Sharpe ratio of 1.1 after transaction costs. Interestingly enough, this tendency is even stronger when focusing on downside deviation. Negative deviations from the mean are substantially lower for PCI at 3.3 percent, compared to MKT with 10.0 percent. In consequence, the Sortino ratio, scaling mean return by unit of downside risk, is at 3.8 for PCI versus 1.0 for MKT - reflecting a strategy with the potential to perform well even in adverse market environments.

Table 12 presents exposures to common sources of systematic risk for the PCI strategy after transaction costs. We perform three regressions, as in Krauss and Stübinger (2015). First, we use the Fama-French three-factor model (FF3), in line with Fama and French (1996). This model measures exposure to the general market, small minus big capitalization stocks (SMB), as well as high minus low book-to-market stocks (HML). Second, we add a

	A: Before transaction costs				B: After transaction costs				
	GGR	CI1	CI2	PCI	GGR	CI1	CI2	PCI	MKT
Mean return	0.0174	0.0223	0.0168	0.1466	0.0111	0.0161	0.0109	0.1234	0.0960
Mean excess return	-0.0119	-0.0070	-0.0124	0.1141	-0.0180	-0.0131	-0.0182	0.0915	0.0647
Standard deviation	0.0331	0.0557	0.0491	0.0845	0.0330	0.0557	0.0488	0.0824	0.1499
Downside deviation	0.0225	0.0347	0.0324	0.0318	0.0233	0.0355	0.0331	0.0328	0.1000
Sharpe ratio	-0.3591	-0.1262	-0.2519	1.3497	-0.5473	-0.2350	-0.3729	1.1099	0.4314
Sortino ratio	0.7718	0.6441	0.5198	4.6077	0.4747	0.4537	0.3277	3.7642	0.9604

Table 11: Annualized return and risk measures of GGR, CI1, CI2, PCI compared to MKT from January 1990 until October 2015. Mean return and mean excess return are calculated on fully invested capital.

momentum and a short-term reversal factor, as in Gatev et al. (2006) and call this variant Fama-French 3+2-factor model (FF3+2). Third, we deploy the recently developed Fama-French five-factor model (FF5), in line with Fama and French (2015). It nests the three-factor model and is enhanced by two additional factors, i.e., portfolios of stocks with robust minus weak profitability (RMW) and with conservative minus aggressive (CMA) investment behavior. The data required for these factor models are downloaded from Kenneth French's website.[6] Across all three models, the PCI strategy results in statistically and economically significant monthly alphas between 0.44 percent and 0.71 percent, after transaction costs. Due to the long-short design of the strategy, exposure to the general market is insignificant and the loading close to zero. The same applies to the SMB, HML, RMW, and CMA factors, with one minor exception. Highest explanatory power has the FF3+2 model, resulting in an R^2 close to ten percent. The increase compared to the other two models is primarily driven by the short-term reversal factor. The loading is positive and highly significant, indicating that the PCI returns can partially be explained by reversal patterns. Conversely, the momentum factor is insignificant, but has the predicted negative sign. Overall, the statistically and economically significant alpha suggests that PCI-based pairs trading is vastly different from basic reversal strategies.

[6] We thank Kenneth R. French for offering all relevant data for these models on his website.

	FF3	FF3+2	FF5
(Intercept)	0.0071***	0.0044**	0.0070***
	(0.0013)	(0.0015)	(0.0014)
Market	0.0511	0.0126	0.0549
	(0.0323)	(0.0333)	(0.0367)
SMB	0.0070	0.0160	
	(0.0433)	(0.0419)	
HML	0.0861	0.0953*	
	(0.0458)	(0.0447)	
Momentum		−0.0150	
		(0.0291)	
Reversal		0.1547***	
		(0.0316)	
SMB5			0.0122
			(0.0485)
HML5			0.0741
			(0.0650)
RMW5			0.0123
			(0.0693)
CMA5			0.0187
			(0.0929)
R^2	0.0164	0.0966	0.0166
Adj. R^2	0.0067	0.0818	0.0005
Num. obs.	310	310	310
RMSE	0.0234	0.0225	0.0234

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 12: Exposure of PCI strategy after transaction costs to systematic sources of risk from January 1990 until October 2015.

4.2.2. Sub-period analysis

Prior pairs trading research suggests that performance fluctuates over time, so we conduct a sub-period analysis (Gatev et al., 2006; Do and Faff, 2010; Bowen and Hutchinson, 2014).

The first sub-period ranges from 01/90 to 03/01. It ends with the introduction of decimalization and the onset of the dot-com crisis. Compared to the overall period presented in table 11, we observe that all strategies perform better in terms of annualized mean returns. PCI still outperforms the other variants by far, with annualized returns of 22 percent after transaction costs and a Sharpe ratio well above 2. Note that during this time period, returns are more likely to be driven by bid-ask bounce, with the minimum price increment set at 1/16 USD. However, at approximately one trade per month, PCI returns roughly 1.5 percent per trade - a value well superior to microstructure effects in a highly liquid stock universe.

The second sub-period ranges from 04/01 until 08/08. It corresponds to a time of moderation. Returns of classical pairs trading strategies decline to zero, or even turn negative. In contrast, PCI still produces 5.2 percent in annualized returns after transaction costs - similar to MKT, but at approximately 50 percent of the standard deviation and only 30 percent of maximum drawdown.

	Before transaction costs				After transaction costs				
Period 01/90-03/01	GGR	CI1	CI2	PCI	GGR	CI1	CI2	PCI	MKT
Mean return	0.0312	0.0486	0.0520	0.2596	0.0245	0.0413	0.0441	0.2231	0.1331
Mean excess return	-0.0180	-0.0013	0.0019	0.2005	-0.0244	-0.0083	-0.0056	0.1656	0.0794
Standard deviation	0.0359	0.0510	0.0537	0.0705	0.0358	0.0510	0.0532	0.0691	0.1476
Sharpe ratio	-0.5020	-0.0261	0.0352	2.8436	-0.6831	-0.1636	-0.1056	2.3963	0.5378
Maximum drawdown	0.1057	0.0622	0.0915	0.0502	0.1138	0.0694	0.0968	0.0526	0.2582
Calmar ratio	0.2953	0.7809	0.5684	5.1658	0.2152	0.5952	0.4560	4.2422	0.5153
Period 04/01-08/08									
Mean return	-0.0001	0.0077	-0.0133	0.0643	-0.0060	0.0024	-0.0177	0.0516	0.0466
Mean excess return	-0.0263	-0.0187	-0.0392	0.0365	-0.0321	-0.0238	-0.0435	0.0241	0.0193
Standard deviation	0.0321	0.0612	0.0406	0.0701	0.0320	0.0613	0.0408	0.0683	0.1343
Sharpe ratio	-0.8188	-0.3059	-0.9651	0.5213	-1.0025	-0.3886	-1.0662	0.3533	0.1437
Maximum drawdown	0.0798	0.2236	0.1297	0.0983	0.1099	0.2430	0.1454	0.0991	0.3218
Calmar ratio	-0.0009	0.0342	-0.1024	0.6539	-0.0548	0.0101	-0.1215	0.5205	0.1449
Period 09/08-12/09									
Mean return	0.0596	0.0338	-0.0253	0.5280	0.0518	0.0290	-0.0309	0.4894	-0.0358
Mean excess return	0.0570	0.0313	-0.0276	0.5245	0.0492	0.0265	-0.0332	0.4859	-0.0383
Standard deviation	0.0465	0.0758	0.0918	0.1889	0.0464	0.0759	0.0916	0.1864	0.2670
Sharpe ratio	1.2264	0.4125	-0.3011	2.7768	1.0604	0.3490	-0.3629	2.6069	-0.1434
Maximum drawdown	0.0375	0.0531	0.0919	0.0306	0.0386	0.0535	0.0939	0.0334	0.3873
Calmar ratio	1.5896	0.6366	-0.2752	17.2665	1.3415	0.5421	-0.3289	14.6515	-0.0924
Period 01/10-10/15									
Mean return	0.0040	-0.0109	-0.0009	-0.0151	-0.0018	-0.0167	-0.0056	-0.0276	0.1308
Mean excess return	0.0036	-0.0113	-0.0013	-0.0154	-0.0022	-0.0171	-0.0060	-0.0280	0.1304
Standard deviation	0.0226	0.0509	0.0303	0.0492	0.0224	0.0508	0.0303	0.0491	0.1386
Sharpe ratio	0.1578	-0.2225	-0.0423	-0.3139	-0.0975	-0.3364	-0.1983	-0.5708	0.9409
Maximum drawdown	0.0379	0.1016	0.0454	0.1297	0.0499	0.1222	0.0501	0.1883	0.1770
Calmar ratio	0.1043	-0.1075	-0.0195	-0.1161	-0.0359	-0.1367	-0.1119	-0.1467	0.7392

Table 13: Annualized risk-return characteristics of GGR, CI1, CI2, PCI compared to MKT for the indicated sub-periods. Mean return and mean excess return are calculated on fully invested capital.

The third sub-period ranges from 09/08 until 12/09 - the global financial crisis. Previous studies have established the fact that pairs trading - as a liquidity providing strategy - performs exceptionally well in low liquidity environments. Specifically, Do and Faff (2010) and Bowen and Hutchinson (2014) find pairs trading to outperform during bear markets, i.e., the dot-com crash or the global financial crisis. We can confirm this finding for almost all our variants. After transaction costs, standard distance-based pairs trading returns more than 5 percent in annualized returns at that time - even better than cointegration-based pairs trading CI1 with close to 3 percent. PCI exhibits the strongest outperformance of the entire sample period with 49 percent in annualized returns, a drawdown of merely 3 percent (compared to 39 percent for MKT), and a Sharpe ratio of 2.6. Given that the strategy buys short-term losers and shorts short-term winners, i.e., provides liquidity, an execution during that time would have been likely. This finding in such recent times posits a severe challenge to the semi-strong form of market efficiency.

The fourth sub-period ranges from 01/10 until 10/15. In light of a market increasing 13 percent on an annualized basis, all strategies disappoint, including PCI. Annualized returns are all close to zero in the negative domain. With active pairs trading research and more than 90 key contributions ever since 2006 (Krauss, 2015), we can only assume that increasing deployment of such strategies renders the markets more efficient.

5. Conclusions

With our paper, we have made three contributions to the literature. The first contribution is conceptual. We have developed the partial cointegration model, established its representation in state space, and provided criteria for identifiability. Also, we have derived a maximum likelihood based estimation routine and a suitable likelihood ratio test. Our model constitutes a novelty, as it allows for the coexistence of a random walk with a mean-reverting overlay. Existing enhancements of classical cointegration models, such as fractional cointegration (Granger, 1986; Granger and Joyeux, 1980; Hosking, 1981) or threshold cointegration (Balke and Fomby, 1997) do not allow for permanent shocks. The same applies to any other advancement of cointegration - at least to our knowledge.

The second contribution is simulative. We find that the maximum likelihood based estimation routine provides consistent estimates as long as the the AR(1) part does not degenerate to a random walk and as long as the proportion of variance attributable to mean-reversion is unequal to zero. Also, the associated likelihood ratio test allows for effective testing against the random walk null. If sample size is sufficiently high, the power is acceptable. Furthermore, we have evaluated PCI-based and cointegration-based pairs trading on artificially generated data. We find the former to outperform the latter for medium to low AR(1)-coefficients, i.e., a short half-life of mean-reversion and for processes with medium to high proportion of variance attributable to mean reversion.

The third contribution is empirical. We deploy partial cointegration in a pairs trading context, where we expect the coexistence of transient components and permanent shocks. We find PCI-based pairs trading to outperform classical cointegration-based and distance-based pairs trading variants. Specifically, we find average annualized returns of more than 12 percent after transaction costs, that can only partially be explained by common sources of systematic risk. Performance is especially strong at times of high market turmoil, and is peaking during the global financial crisis with annualized returns of 49 percent. Only in recent years, performance has leveled off, suggesting that the markets may have reacted to recent advancements in pairs trading research. Nevertheless, our findings pose a severe challenge to the semi-strong form of market efficiency, considering the consistently high returns PCI-based pairs trading produces from 1990 until 2010 in a highly-liquid stock universe.

For further research, we see the potential to deploy this model in other economic contexts, where the coexistence of transient and permanent shocks could be expected. For example, future markets (spreads between spot and future prices), commodities (spread between futures on raw products and refined end products) or bond markets (spread between short-term and long-term debt) - to name a few. Also, classical macroeconomic phenomena typical for cointegration analysis, such as income (Campbell, 1987), money demand (Johansen and Juselius, 1990), or purchasing power parity (Corbae and Ouliaris, 1988) could be reinvestigated with partial cointegration.

AppendixA. Identifiability

A system is said to be identifiable if the parameters of the system can be uniquely determined from a (possibly infinite) realization of that system. In this section, we show that if X_1 and X_2 conform to the model given by equation (1), then the system can be identified if the following conditions are met:

1. The sequence of first differences $((1 - B)X_{1,t})_{t \in T}$ exhibits weak stationarity, and
2. The sequence of first differences $((1 - B)X_{1,t})_{t \in T}$ is independent of the sequence of first differences $((1 - B)(M_t + R_t))_{t \in T}$.

To see that a system is identifiable in case assumptions 1 and 2 are fulfilled, consider the state space representation given in equations (7), (8). Then, the first difference of $X_{1,t}$ is $\varepsilon_{X,t}$, with variance σ_X^2, which is assumed to be independent of $\varepsilon_{M,t}$ and $\varepsilon_{R,t}$. Now, let $D_t = (1 - B)X_{1,t}$. By assumption, $\mathrm{Var}[D_t]$ is defined, and D_t is independent of $(1 - B)(M_t + R_t)$, and therefore $\mathrm{Cov}[D_t, (1 - B)(M_t + R_t)] = 0$. Also, note that

$$(1 - B)X_{2,t} = (1 - B)(\beta X_{1,t} + M_t + R_t) = \beta D_t + (1 - B)(M_t + R_t). \tag{A.1}$$

Consequently,

$$\begin{aligned}
\mathrm{Cov}[(1 - B)X_{1,t}, (1 - B)X_{2,t}] &= \mathrm{Cov}[D_t, \beta D_t + (1 - B)(M_t + R_t)] && \text{(A.2)} \\
&= \mathrm{Cov}[D_t, \beta D_t] + \mathrm{Cov}[D_t, (1 - B)(M_t + R_t)] && \text{(A.3)} \\
&= \beta \mathrm{Var}[D_t]. && \text{(A.4)}
\end{aligned}$$

Therefore, β can be recovered as

$$\beta = \frac{\mathrm{Cov}[(1 - B)X_{1,t}, (1 - B)X_{2,t}]}{\mathrm{Var}[(1 - B)X_{1,t}]}. \tag{A.5}$$

Having recovered β, we can now compute the sequence

$$W_t = X_{2,t} - \beta X_{1,t} = M_t + R_t, \tag{A.6}$$

which is partially autoregressive, and known to be identifiable (Clegg, 2015). If we take $v_k = \mathrm{Var}[(1-B^k)W_t]$, the remaining parameters can be determined following Clegg (2015):

$$\rho = -\frac{v_1 - 2v_2 + v3}{2v_1 - v_2} \tag{A.7}$$

$$\sigma_M^2 = \frac{1}{2}\left(\frac{\rho+1}{\rho-1}\right)(v_2 - 2v_1) \tag{A.8}$$

$$\sigma_R^2 = \frac{1}{2}\left(v_2 - 2\sigma_M^2\right). \tag{A.9}$$

AppendixB. Likelihood function

Let $\boldsymbol{\Theta_t}$ denote the information that was available up to and including time t, and let $\mathbf{\Phi}$ denote the parameter values $\beta, \rho, \sigma_X, \sigma_M$, and σ_R. Given the sequence of observations $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ and parameter values $\mathbf{\Phi}$, the likelihood function can be written as

$$\mathcal{L}(\boldsymbol{\phi}) = p(\boldsymbol{X_1}|\mathbf{\Phi}) \prod_{k=2}^{n} p(\boldsymbol{X_k}|\boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}). \tag{B.1}$$

Using the Markov property, this can be rewritten as

$$\mathcal{L}(\boldsymbol{\phi}) = p(\boldsymbol{X_1}|\mathbf{\Phi}) \prod_{k=2}^{n} p(\boldsymbol{X_k}|\boldsymbol{X_{k-1}}, \mathbf{\Phi}). \tag{B.2}$$

Expanding this in terms of the definition of $\boldsymbol{X_t}$, we have

$$\mathcal{L}(\boldsymbol{\phi}) = p(X_{1,1}, X_{2,1}|\mathbf{\Phi}) \prod_{k=2}^{n} p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}). \tag{B.3}$$

Now, we focus on a particular term in this product, $p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi})$. From the laws of conditional probability, we have $p(A, B|C) = p(A|B, C)p(B|C)$. Therefore,

$$p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}) = p(X_{2,k}|X_{1,k}, X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi})p(X_{1,k}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}). \tag{B.4}$$

We proceed by separately evaluating each of the two terms on the right hand side of the above equation. To start with, we have

$$\begin{aligned} p(X_{1,k}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}) &= p(X_{1,k} - X_{1,k-1}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}) && \text{(B.5)} \\ &= p(\varepsilon_{X,k}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}) && \text{(B.6)} \\ &= \phi(\varepsilon_{X,k}; 0, \sigma_X^2), && \text{(B.7)} \end{aligned}$$

where $\phi(\cdot)$ denotes the probability density function of the normal distribution - in this case with expectation zero and variance σ_X^2.

To evaluate $p(X_{2,k}|X_{1,k}, X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi})$, we begin by noting that

$$X_{2,k} - E(X_{2,k}|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) = \beta X_{1,k} + M_k + R_k - E(\beta X_{1,k} + M_k + R_k|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) \tag{B.8}$$

$$= M_k + R_k - E(M_k + R_k|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) \tag{B.9}$$

$$= M_k - E(M_k|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) + R_k - E(R_k|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) \tag{B.10}$$

$$= \rho M_{k-1} + \varepsilon_{M,k} - E(\rho M_{k-1} + \varepsilon_{M,k}|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) + \tag{B.11}$$

$$R_{k-1} + \varepsilon_{R,k} - E(R_{k-1} + \varepsilon_{R,k}|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) \tag{B.12}$$

$$= \varepsilon_{M,k} - E(\varepsilon_{M,k}|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) + \tag{B.13}$$

$$\varepsilon_{R,k} - E(\varepsilon_{R,k}|X_{1,k}, \boldsymbol{\Theta_{k-1}}, \mathbf{\Phi}) \tag{B.14}$$

$$= \varepsilon_{M,k} + \varepsilon_{R,k}. \tag{B.15}$$

Therefore,

$$p(X_{2,k}|X_{1,k}, X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}) = p(\varepsilon_{M,k} + \varepsilon_{R,k}|X_{1,k}, X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}) \tag{B.16}$$

$$= \phi(\varepsilon_{M,k} + \varepsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2). \tag{B.17}$$

Putting the above together and returning to the likelihood function, we therefore have

$$\mathcal{L}(\boldsymbol{\phi}) = p(X_{1,1}, X_{2,1}; \mathbf{\Phi}) \prod_{k=2}^{n} p(X_{1,k}, X_{2,k}|X_{1,k-1}, X_{2,k-1}, \mathbf{\Phi}) \tag{B.18}$$

$$= p(X_{1,1}, X_{2,1}; \mathbf{\Phi}) \prod_{k=2}^{n} \phi(\varepsilon_{X,k}; 0, \sigma_X^2)\phi(\varepsilon_{M,k} + \varepsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2) \tag{B.19}$$

$$= p(X_{1,1}, X_{2,1}; \mathbf{\Phi}) \Big(\prod_{k=2}^{n} \phi(\varepsilon_{X,k}; 0, \sigma_X^2) \Big) \Big(\prod_{k=2}^{n} \phi(\varepsilon_{M,k} + \varepsilon_{R,k}; 0, \sigma_M^2 + \sigma_R^2) \Big) \tag{B.20}$$

$$= p(X_{1,1}, X_{2,1}; \mathbf{\Phi}) \mathcal{L}_X(\sigma_X) \mathcal{L}_{MR}(\beta, \rho, \sigma_M, \sigma_R). \tag{B.21}$$

AppendixC. Likelihood ratio test

Let

$$\mathcal{L}_{MR}^* = \max_{\beta,\rho,\sigma_M,\sigma_R} \mathcal{L}_{MR}(\beta,\rho,\sigma_M,\sigma_R) \tag{C.1}$$

be the maximum value of the likelihood function found by the optimization routine outlined in 2.2, and similarly let

$$\mathcal{L}_{RW}^* = \max_{\beta,\sigma_R} \mathcal{L}_{MR}(\beta,\rho=0,\sigma_M=0,\sigma_R) \tag{C.2}$$

be the maximum value of this function that is found when the parameters ρ and σ_M are held constant at zero. Then, the value of the test statistic is given as

$$\Lambda = \log\left(\frac{\mathcal{L}_{RW}^*}{\mathcal{L}_{MR}^*}\right) \tag{C.3}$$

Critical values for testing the simplified null hypothesis $\mathcal{H}_0^R$ that the cointegrating process is a random walk are determined through simulation. Specifically, 1000 random partially cointegrated pairs are generated with the parameters $\beta = 1, \sigma_X = 1, \sigma_M = 0$, and $\sigma_R = 1$. For each such pair, the log likelihood ratio score Λ is computed, and the quantiles are tabulated. This procedure is repeated for various values of the sample size n.

n	p=0.01	p=0.05	p=0.10
50	-5.3	-3.3	-2.4
100	-5.4	-3.3	-2.4
250	-5.1	-3.0	-2.2
500	-4.7	-3.1	-2.3
1000	-4.6	-3.1	-2.4
2500	-5.0	-3.3	-2.5

Table C.14: Critical values for the likelihood ratio test for the null hypothesis that the cointegrating process is a random walk.

Bibliography

Avellaneda, M., Lee, J.-H., 2010. Statistical arbitrage in the US equities market. Quantitative Finance 10 (7), 761–782.

Bacon, C. R., 2008. Practical portfolio performance: Measurement and attribution, 2nd Edition. Wiley finance. Wiley, Chichester, England and Hoboken, NJ.

Balke, N. S., Fomby, T. B., 1997. Threshold Cointegration. International Economic Review 38 (3), 627–645.

Bertram, W. K., 2010a. A non-stationary model for statistical arbitrage trading. URL http://ssrn.com/abstract=1632718

Bertram, W. K., 2010b. Analytic solutions for optimal statistical arbitrage trading. Physica A: Statistical Mechanics and its Applications 389 (11), 2234–2243.

Bogomolov, T., 2013. Pairs trading based on statistical variability of the spread process. Quantitative Finance 13 (9), 1411–1430.

Bowen, D. A., Hutchinson, M. C., 2014. Pairs trading in the UK equity market: Risk and return. The European Journal of Finance, 1–25.

Brockwell, P. J., Davis, R. A., 2010. Introduction to time series and forecasting, second edition Edition. Springer texts in statistics. Springer, New York.

Brogaard, J., Hendershott, T., Riordan, R., 2014. High-frequency trading and price discovery. Review of Financial Studies 27 (8), 2267–2306.

Caldeira, J. F., Moura, G. V., 2013. Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. Brazilian Review of Finance 11 (1), 49–80.

Campbell, J. Y., 1987. Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis. Econometrica 55 (6), 1249–1274.

Caner, M., Hansen, B. E., 2001. Threshold autoregression with a unit root. Econometrica 69 (6), 1555–1596.

Clegg, M., 2014. On the persistence of cointegration in pairs trading. SSRN Electronic Journal.
URL http://ssrn.com/abstract=2491201

Clegg, M., 2015. Modeling time series with both permanent and transient components using the partially autoregressive model. SSRN Electronic Journal.
URL http://ssrn.com/abstract=2556957

Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance 1 (2), 223–236.

Corbae, D., Ouliaris, S., 1988. Cointegration and tests of purchasing power parity. The Review of Economics and Statistics, 508–511.

de Moura, C. E., Pizzinga, A., Zubelli, J., 2016. A pairs trading strategy based on linear state space models and the Kalman filter. Quantitative Finance, 1–15.
URL http://dx.doi.org/10.1080/14697688.2016.1164886

Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74 (366), 427–431.

Do, B., Faff, R., 2010. Does simple pairs trading still work? Financial Analysts Journal 66 (4), 83–95.

Do, B., Faff, R., 2012. Are pairs trading profits robust to trading costs? Journal of Financial Research 35 (2), 261–287.

Do, B., Faff, R., Hamza, K., 2006. A new approach to modeling and estimation for pairs trading. In: Proceedings of 2006 Financial Management Association European Conference.

Dunis, C. L., Giorgioni, G., Laws, J., Rudy, J., 2010. Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities. Working paper, Liverpool Business School.

Durbin, J., Koopman, S. J., 2012. Time series analysis by state space methods, second edition Edition. Vol. 38 of Oxford statistical science series. Oxford University Press, Oxford.

Elliott, R. J., Van Der Hoek*, John, Malcolm, W. P., 2005. Pairs trading. Quantitative Finance 5 (3), 271–276.

Engle, R. F., Granger, C. W. J., 1987. Co-Integration and error correction: Representation, estimation, and testing. Econometrica 55 (2), 251.

Fama, E. F., French, K. R., 1996. Multifactor explanations of asset pricing anomalies. The Journal of Finance 51 (1), 55.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116 (1), 1–22.

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 1999. Pairs trading: Performance of a relative value arbitrage rule. Working paper, Yale School of Management's International Center for Finance.

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies 19 (3), 797–827.

Granger, C. W., 1986. Developments in the study of cointegrated economic variables. Oxford Bulletin of Economics and Statistics 48 (3), 213–228.

Granger, C. W. J., Joyeux, R., 1980. An introduction to long–memory time series models and fractional differencing. Journal of Time Series Analysis 1 (1), 15–29.

Harbo, I., Johansen, S., Nielsen, B., Rahbek, A., 1998. Asymptotic inference on cointegrating rank in partial systems. Journal of Business & Economic Statistics 16 (4), 388–399.

Hosking, J. R. M., 1981. Fractional differencing. Biometrika 68 (1), 165–176.

Huck, N., 2015. Pairs trading: Does volatility timing matter? Applied Economics 47 (57), 6239–6256.

Huck, N., Afawubo, K., 2015. Pairs trading and selection methods: is cointegration superior? Applied Economics 47 (6), 599–613.

Ince, O. S., Porter, B. R., 2006. Individual equity return data from Thomson Datastream: Handle with care! Journal of Financial Research 29 (4), 463–479.

Jacobs, H., Weber, M., 2015. On the determinants of pairs trading profitability. Journal of Financial Markets 23, 75–97.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12 (2-3), 231–254.

Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration–with applications to the demand for money. Oxford Bulletin of Economics and Statistics 52 (2), 169–210.

Jondeau, E., Lahaye, J., Rockinger, M., 2015. Asymmetry in the price impact of trades in an high-frequency microstructure model with jumps. Journal of Banking and Finance, forthcoming.

Jones, C. M., 2002. A century of stock market liquidity and trading costs. Working paper, Columbia University.

Krauss, C., 2015. Statistical arbitrage pairs trading strategies: Review and outlook. IWQW Discussion Paper Series, University of Erlangen-Nürnberg.

Krauss, C., Do, X. A., Huck, N., 2016. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. FAU Discussion Papers in Economics, University of Erlangen-Nürnberg.

Krauss, C., Herrmann, K., Teis, S., 2015. On the power and size properties of cointegration tests in the light of high-frequency stylized facts. IWQW Discussion Paper Series, University of Erlangen-Nürnberg.

Krauss, C., Stübinger, J., 2015. Nonlinear dependence modeling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100. IWQW Discussion Paper Series, University of Erlangen-Nürnberg.

Krishnakumar, J., Neto, D., 2009. Estimation and testing for the cointegration rank in a threshold cointegrated system. Working paper, Cahiers du département d'économétrie, Faculté des sciences économiques et sociales, Université de Genève.

Leippold, M., Lohre, H., 2012. International price and earnings momentum. The European Journal of Finance 18 (6), 535–573.

Nath, P., 2003. High frequency pairs trading with U.S. treasury securities: Risks and rewards for hedge funds. Working paper, London Business School.

Pantula, S. G., Gonzalez-Farias, G., Fuller, W. A., 1994. A comparison of unit-root test criteria. Journal of Business & Economic Statistics 12 (4), 449–459.

Peterson, B. G., Carl, P., 2014. PerformanceAnalytics: Econometric tools for performance and risk analysis. R package.
URL http://CRAN.R-project.org/package=PerformanceAnalytics

Phillips, P. C. B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75 (2), 335–346.

Poterba, J. M., Summers, L. H., 1988. Mean reversion in stock prices. Journal of Financial Economics 22 (1), 27–59.

Prager, R., Vedbrat, S., Vogel, C., Watt, E. C., 2012. Got liquidity? BlackRock Investment Institute.

Rad, H., Low, Rand Kwong Yew, Faff, R. W., 2015. The profitability of pairs trading strategies: Distance, cointegration, and copula Methods. Working paper, University of Queensland.

Simon, D., 2006. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. John Wiley & Sons, Hoboken, N.J.

S&P Dow Jones Indices, 2015. Equity S&P 500.
URL http://us.spindices.com/indices/equity/sp-500

Summers, L. H., 1986. Does the stock market rationally reflect fundamental values? The Journal of Finance 41 (3), 591–601.

Triantafyllopoulos, K., Montana, G., 2011. Dynamic modeling of mean-reverting spreads for statistical arbitrage. Computational Management Science 8 (1-2), 23–49.

Vidyamurthy, G., 2004. Pairs trading: Quantitative methods and analysis. John Wiley & Sons, Hoboken, N.J.