Ultra-high-frequency pairs trading in gold ETFs

Thong M Dao, Centre for Digital Finance, Southampton Business School, University of Southampton, SO17 1BJ, UK, tmd2g13@soton.ac.uk

Frank McGroarty, Centre for Digital Finance, Southampton Business School, University of Southampton, SO17 1BJ, UK, f.j.mcgroarty@soton.ac.uk

Andrew Urquhart, Centre for Digital Finance, Southampton Business School, University of Southampton, SO17 1BJ, UK, aju1y12@soton.ac.uk

Abstract

Based on a large dataset of gold ETFs, we find arbitrage opportunities in the gold ETF market which can be exploited by high-frequency traders. To our knowledge, this is the first paper to study pairs trading of gold ETFs using tick data. Able to execute their orders with minimal delay and take advantage of potentially short-lived opportunities, high-frequency traders can make an excess return of 2.1% p.a. after including transaction costs. Consistent with Grossman and Stiglitz (1976) and Grossman and Stiglitz (1980), this profitability may be compensation for arbitrage efforts and incentivise arbitrageurs to eliminate mispricing. We also explain why the trade exit rule of full convergence used in previous studies may not be optimal and propose a rule based on partial convergence which outperforms the standard full-convergence rule. Specifically, changing the exit rule from full convergence to partial convergence can increase pairs trading returns to 3.38% p.a. and also enhance the risk-adjusted performance based on different risk adjustment methods. More importantly, the outperformance of partial convergence is consistent in both the whole sample and the sub-samples with the optimal convergence target being around 40%. Therefore, partial convergence enables better exploitation of arbitrage opportunities than full convergence. Finally, the pairs trading returns exceed compensation for risks, which suggests that the gold ETF market may be inefficient at ultra-high frequency.

Keywords: pairs trading, statistical arbitrage, high frequency, gold ETFs

1. Introduction

Arbitrage and the law of one price are among the most important and extensively studied principles in finance. Many papers have documented profitable mispricing (e.g. Froot and Dabora, 1999, Gatev et al., 2006, Gagnon and Karolyi, 2010, Do and Faff, 2012, Marshall et al., 2013) whereas many others have emphasised limits to arbitrage (e.g. De Long et al., 1990, Shleifer and Vishny, 1997, Abreu and Brunnermeier, 2002, D'Avolio, 2002, Grossmann et al., 2007). One particular arbitrage strategy, known as statistical arbitrage or pairs trading, is widely used by hedge funds and investment banks. The idea of pairs trading is to find two closely related stocks which often move together, wait for their temporary price divergence (i.e. relative mispricing) and exploit the subsequent convergence by buying the underpriced stock and selling the overpriced one simultaneously (Gatev et al., 2006). Pairs trading works because similar instruments in imperfectly efficient markets may diverge in price but this disequilibrium will be arbitraged away (Do and Faff, 2010).

Pairs trading can help maintain market efficiency because arbitrage activities help reduce the mispricings between instruments (Kondor, 2009). Alsayed and McGroarty (2012) examine the law of one price in the ADR [1] market and show that pairs trading plays a major role in maintaining the stock-ADR price parity. These findings are consistent with the argument of Grossman and Stiglitz (1980) that (i) markets are not perfectly and automatically efficient, (ii) there must be some inefficiencies so that traders are motivated to trade and exploit them and (iii) these trading activities reduce market inefficiencies and make markets more efficient.

Our motivation is the profitability of pairs trading documented in the literature (e.g. Gatev et al., 2006, Do and Faff, 2010, Do and Faff, 2012, Jacobs and Weber, 2015). In the seminal paper on pairs trading, Gatev et al. (2006) find that this strategy can generate a return of 11% p.a. while recently, Jacobs and Weber (2015) find that pairs trading can generate at least 12% p.a. consistently. However, pairs trading based on ultra-high-frequency data has not received much attention so we contribute to the literature by examining pairs trading using tick data.

The speed of information gathering and action taking in financial markets is increasingly fast (see Goldstein et al., 2014 for an overview of high-frequency trading). In today's market, being marginally faster than other competitors can make a significant difference to trading results,

[1] ADR stands for American Depositary Receipt.

especially during volatile periods (e.g. news announcements). As a result, traders have been doing whatever it takes to gain a speed advantage, from investing heavily in technology to co-locating their trading systems inside the trading venue. Hasbrouck and Saar (2013) find that this arms race has reduced the reaction time of traders to only a few milliseconds, hundreds of times faster than the time it takes to perform an eye blink. High-frequency traders generate most trading activities and thus are important participants in the markets. We find that pairs trading is profitable for high-frequency traders. The pairs trading excess returns (2.1% per annum) may encourage arbitrage activities and compensate for arbitrage risks and costs (Grossman and Stiglitz, 1976, Grossman and Stiglitz, 1980).

Most pairs trading studies (e.g. Gatev et al., 2006, Do and Faff, 2010, Do and Faff, 2012) rely on full convergence (i.e. the two stocks converge completely after their temporary divergence). We explain why the trade exit rule of full convergence may not be optimal and may lead to underestimation of pairs trading profitability and the level of market inefficiency. Therefore, we propose a better rule based on partial convergence to improve the pairs trading mechanism. Partial convergence is achieved when the two stocks converge to some extent but not completely and thus it is a more flexible condition than full convergence. We show that the partial-convergence rule outperforms the full-convergence rule in both the whole sample and the sub-samples and thus enables better exploitation of trading opportunities and more accurate reflection of market inefficiency.

Our study on high-frequency pairs trading is conducted in the context of the fast growing ETF market (Shin and Soydemir, 2010, Caginalp et al., 2014, Kearney et al., 2014). Our choice is motivated by findings of ETF mispricing in the literature (e.g. Ackert and Tian, 2000, Engle and Sarkar, 2006). If there are mispricings between ETFs, the ETF market is an interesting and potentially profitable environment for pairs trading which exploits relative mispricing. Moreover, this setting is appropriate for an arbitrage study because ETFs are easily accessible to traders and less risky to arbitrage than in other settings (Marshall et al., 2013). Specifically, divergence risk (i.e. the two stocks do not converge or they take a long time to do so) is low since (i) investors buy ETFs to track the underlying index or asset so the ETFs' management must try to minimise tracking errors to attract investors and (ii) ETF shares can be exchanged for the underlying asset or component stocks to benefit from potential mispricing, which should keep prices in equilibrium (Engle and Sarkar, 2006).

Among different types of ETFs, we focus on gold ETFs. As a result of the significant appreciation of gold in the first decade after 2000 (Pullen et al., 2014), gold investment and attention to gold in the literature have been growing (see O'Connor et al., 2015 for an overview). At only $250 per ounce in 2001, the gold price has increased to over $1500 per ounce in 2012 (Blose and Gondhalekar, 2013) and gold has remained an important asset to investors, especially in tough times because of its safe-haven properties (Baur and Lucey, 2010, Baur and McDermott, 2010, Bredin et al., 2015). The World Gold Council also reported a substantial increase in the investment demand for gold in general and gold ETFs in particular from July 2008 to March 2009, which has continued to grow (Pullen et al., 2014). The daily global turnover of gold was estimated to be 4000 metric tons with a trading value comparable to that of all stock exchanges in the world combined [2]. Gold is often considered a part of the currency market, the largest market in the world, and its turnover is lower than only four currency pairs (Hauptfleisch et al., 2016). Given the importance of gold and gold ETFs, our study on pairs trading of gold ETFs may have valuable implications for investors. We hope to provide insights into ETF pairs trading beyond the equity ETFs in Marshall et al. (2013). Because (i) pairs trading exploits the mispricing of ETFs which relates to their ability to track the underlying index or asset and (ii) this tracking ability differs among ETFs, pairs trading performance may differ between different types of ETFs. Indeed, we show that gold ETFs may be inefficient while Marshall et al. (2013) point out that their results do not necessarily support market inefficiency for equity ETFs.

We collect bid-ask quotes (time-stamped to milliseconds) of the two most liquid US gold ETFs, namely SPDR Gold Shares and iShares Gold Trust from 2005 to 2010. We believe this is the first study to examine pairs trading of gold ETFs using tick data and we contribute to the literature on pairs trading in three ways. Firstly, the use of ultra-high frequency data allows us to capture short-lived mispricings unobservable in low frequency data (e.g. daily) which is the data frequency often used in the literature. In fact, while we find very few arbitrage opportunities in the daily data of our gold ETFs, there are a number of opportunities in the tick data. Following Marshall et al. (2013), we only consider relative mispricing of at least 0.2% and find a comparable number of arbitrage opportunities as in equity ETFs found in Marshall et al. (2013) on an annual basis. Secondly, we enhance the standard pairs trading mechanism

[2] Bank of International Settlement's report on the global foreign exchange market activity. Available at http://www.bis.org/publ/rpfxf10t.pdf

used in the literature by proposing a better trading rule based on partial convergence which is supported by both ex-ante justification and ex-post results. Thirdly, we show that the pairs trading returns exceed compensation for risks using different risk adjustment methods, which suggests that the gold ETF market may be inefficient.

This study proceeds as follows. In section 2, we review the literature on pairs trading performance and the explanation for such performance, and propose a classification scheme for pairs trading studies. Section 3 describes our data. Section 4 presents the methodology of our pairs trading analysis based on both full convergence and partial convergence. Section 5 reports the results. Section 6 discusses the results and provides conclusions.

2. Literature review

2.1. Pairs trading performance

Gatev et al. (2006), testing pairs trading in four decades (1962 – 2002) with daily data, find that self-financing pairs trades generate up to 11% of annualised excess returns on average. Pairs trading profits are robust to transaction costs, short-selling costs and short recalls even in out-of-sample test. They also find that a large part of profits comes from short positions. However, their zero-investment assumption (i.e. short sale proceeds are used to finance the long position in a trade so no upfront capital is required) may not be realistic and hence some authors assume a 50% margin requirement (e.g. Mitchell et al., 2002, Marshall et al., 2013). In another paper, Alsayed and McGroarty (2012) indicate that arbitrage profitability does not correlate with broad market performance.

Do and Faff (2010), extending the study of Gatev et al. (2006) to 2009, confirm that although pairs trading performed well before 1990 (even during the 1987 crash), it has deteriorated since the 1990s (evidenced by regular unprofitable months). This deterioration may be caused by the reduced mispricing frequency in recent periods as a result of increasingly popular algorithmic trading (Akram et al., 2009). Nevertheless, Marshall et al. (2013), who study high-frequency pairs trading between US ETFs tracking the S&P500 index from 2001 to 2010, report greater profit magnitude from recent mispricings. Moreover, despite decreasing profitability, pairs trading shows favourable performance in tough times (Do and Faff, 2010) and more importantly, its risk-adjusted returns remain stable over time (Gatev et al., 2006).

Trade duration depends on the timeframe of pairs trading and data frequency. At one end, the average duration of trades on the daily timeframe is 3.75 months (Gatev et al., 2006). At the other end, on the tick timeframe, the strong mean reversion of pairs reduces the arbitrage length to minutes (Alsayed and McGroarty, 2012, Marshall et al., 2013).

2.2. Explanation for pairs trading profitability

The documented profitability of pairs trading results from the relationship between the two stocks in a pair, compensation for arbitrage efforts and microstructure effects. The relationship in a pair refers to both fundamental and technical relationship. Fundamentally, the two pair components are close substitutes when they belong to the same industry (Gatev et al., 2006). Do and Faff (2010) find that industry homogeneity has statistically significant impacts on pairs trading returns and increased granularity of industry classification can improve performance. Specifically, the 48-industry categorisation of Fama and French (1997) performs better than the four-group classification (i.e. Financials, Industrials, Transportation and Utilities). There is also evidence of industry-specific profitability (i.e. Financials and Utilities pairs outperform Industrials and Transportation pairs) because of industry-specific levels of company homogeneity.

The technical relationship refers to price behaviours of the two stocks in a pair. Marshall et al. (2013) find that price divergences are often followed by fast convergences. If the initial divergences are considered overreaction, the subsequent convergences might be reversals of overreaction. Furthermore, the past convergence frequency influences returns. Stocks whose prices often intersect in the formation period (i.e. when pairs are selected based on minimum sum of squared differences between normalised price series) are likely to be profitable pairs in the testing period. Adding the number of price intersections to pairs selection criteria increases mean excess returns (Do and Faff, 2010).

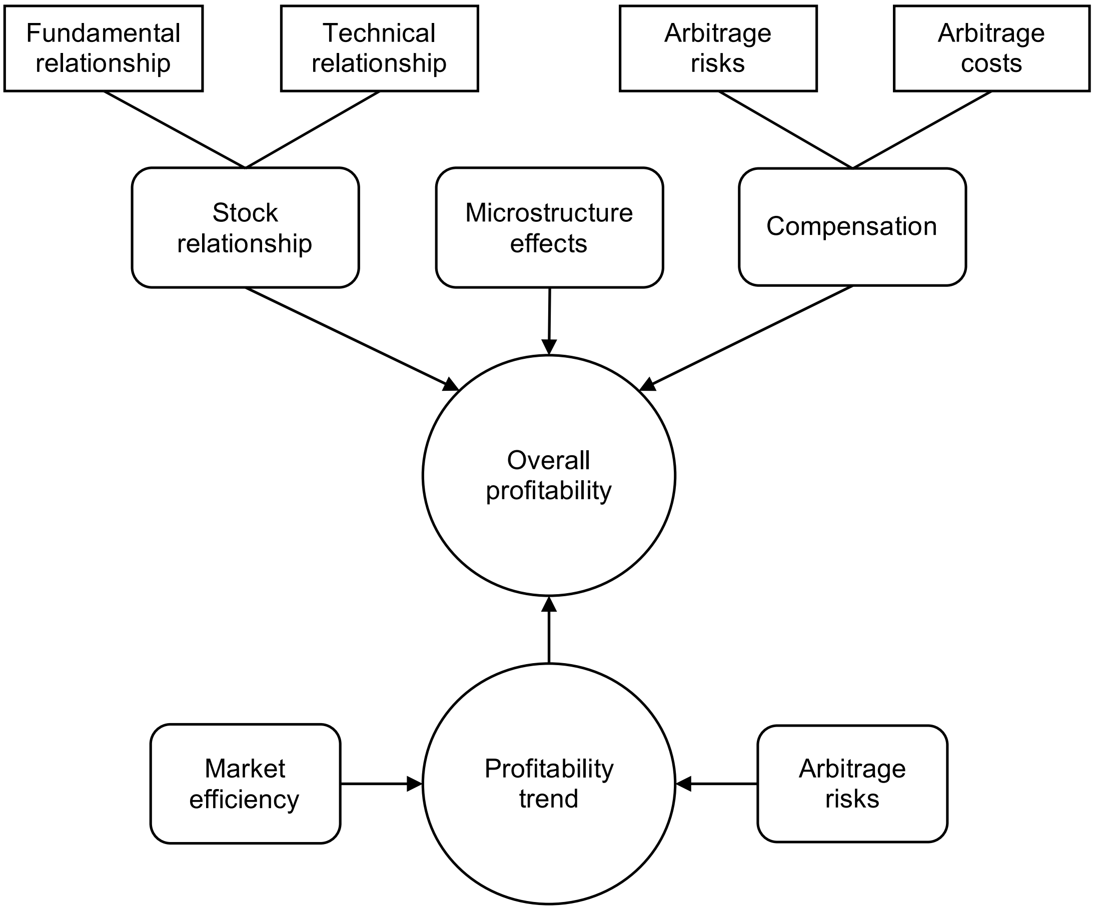
From another viewpoint, mispricing incentivises the enforcement of the Law of One Price [3] (Alsayed and McGroarty, 2012) and pairs trading profits may compensate for non-convergence risk, information cost and other risks and costs (Marshall et al., 2013). Nevertheless, only part

[3] The Law of One Price implies that in efficient markets, financial instruments with the same cash flows should trade at the same price, regardless of their creation methods (Akram et al., 2009). More generally, two securities whose pay-offs are close to each other should have prices which are equally close to each other (Chen and Knez, 1995).

of the profitability is explained by exposure to five factors (i.e. market risk, firm size, firm value, momentum and reversal). Additionally, firm-specific volatility affects arbitrage performance whilst systematic volatility does not (Do and Faff, 2010). Finally, due to the contrarian nature of pairs trading, microstructure effects might bias its observed profitability upward (Conrad and Kaul, 1989, Jegadeesh and Titman, 1995). If trade prices are used instead of quotes, contrarian trades are likely to be taken on the wrong (and more favourable) side of the spread. If the current trend is up (down), the observed trade price is more likely to be at the ask (bid) so a contrarian strategy will mistakenly sell at the ask and buy at the bid. We use quote data to address this issue.

With regards to the downward trend in pairs trading profitability over time, Do and Faff (2010) conclude that it is attributable to an increase in market efficiency and arbitrage risks. To reach this conclusion, they use winning trades and losing trades to capture the market efficiency effect and arbitrage risk effect respectively. If the efficiency effect is significant, profits will be smaller and less frequent over time; if the risk effect is significant, losses will be increasingly large and regular. Although decreasing transaction costs have increased market efficiency by attracting more pairs trading, especially from hedge funds since 1989 (Gatev et al., 2006); only 30% of the performance deterioration is because of improved efficiency while the rest is caused by higher arbitrage risks (Do and Faff, 2010). Figure 1 demonstrates the profitability explanation.

Figure 1. Profitability explanation. This figure explains pairs trading profitability, in general and over time. The stock relationship includes fundamental and technical relationship. The compensation covers arbitrage risks and costs. The arrows show the direction of explanation (i.e. the causes point to the effects). 'Profitability trend' points to 'overall profitability' because the trend is a part of the overall performance.



2.3. Pairs trading classification

Based on the strength of the fundamental relationship between the two stocks in a pair, we propose classifying pairs trading studies into three types, namely the loose form, semi-strict form and strict form. In our loose form, the relationship is purely statistical and the two stocks have no fundamental reason to move together. For example, two stocks in two unrelated industries (e.g. Starbucks and IBM) are likely to be influenced by different factors so they may behave differently and their observed relationship is only statistical. In our semi-strict form, the two stocks belong to the same industry or economic sector and they are likely to move together because both of them are affected by common factors (e.g. industry-specific regulation, supply and demand). For instance, if the demand for banking products is increasing and there are two reasonably comparable banks, it is expected that the demand for both banks will increase. As a result, their stocks should be affected in a similar manner and show similar movements. From an economic viewpoint, the products of the two firms in a semi-strict pair can be substitutes for each other to some extent. Finally, in our strict form, the two stocks have an inherent relationship which should force them to move together if markets are efficient. Specifically, they represent the same index or asset; in other words, they are different covers of the same content and close substitutes for each other (e.g. ETFs tracking the S&P 500 index).

As we move from the loose form to semi-strict form to strict form, the requirement becomes increasingly stringent and the number of eligible pairs decreases. Our three types of pairs trading have a nested relationship as shown in Figure 2. Loose-form pairs include semi-strict-form pairs because pairs in the same industry (i.e. semi-strict form) constitute a subset of all pairs in the economy (i.e. loose form); and similarly, semi-strict-form pairs include strict-form pairs. Some examples of loose-form and semi-strict-form studies are Gatev et al. (2006) and Do and Faff (2010) where they match pairs of US stocks based on statistical criteria at first and then add the criterion of industry homogeneity by matching only stocks in the same category (i.e. Financials, Industrials, Transportation or Utilities). Strict-form studies include Alsayed and McGroarty (2012) and Marshall et al. (2013) who examine UK stock-ADR pairs and a US ETF pair tracking the S&P500 index respectively. We classify the literature in Table 1. Based on our classification, our research on gold ETFs belongs to the strict form because these ETFs track the same asset so they have an inherent relationship.

Figure 2. Three types of pairs trading. This figure shows our three types of pairs trading and their nested relationship. Google-Ford pair belongs to the loose form since they operate in two unrelated industries (i.e. Internet information and automobile). Google-Yahoo pair belongs to the semi-strict form as they are in the same industry and offer fairly similar products and services.

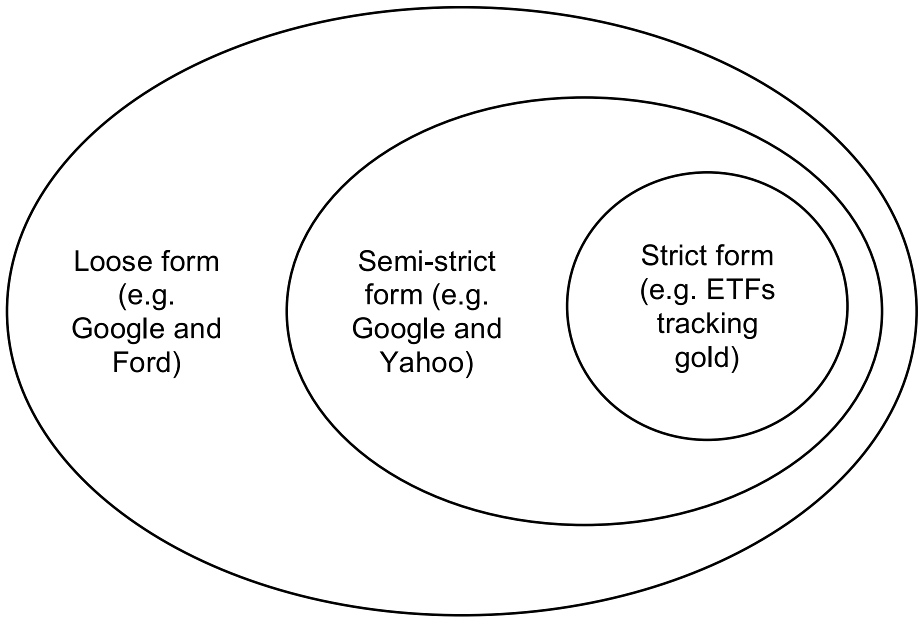


Table 1. Our classification of the literature on pairs trading. Some authors study more than one type of pairs trading.

Type	Author	Market	Period	Data Type	Data Frequency	Findings
Loose form	Gatev et al. (2006)	US	1962 – 2002	Transaction price	Daily	It is possible to make a return of 11% p.a. from self-financing pairs trades.
	Do and Faff (2010)	US	1962 – 2009	Transaction price	Daily	Pairs trading has become less profitable since the 1990s.
	Do and Faff (2012)	US	1963 – 2009	Transaction price	Daily	Pairs trading is slightly profitable after considering trading costs.
	Broussard and Vaihekoski (2012)	Finland	1987 – 2008	Transaction price	Daily	The annualised return from pairs trading can be up to 12.5% and the profits do not relate to systematic risks.
	Jacobs and Weber (2015)	34 countries	2000 – 2013	Transaction price	Daily	Pairs trading is consistently profitable, generating more than 12% p.a..
Semi-strict form	Gatev et al. (2006)	US industry groups	1962 – 2002	Transaction price	Daily	Pairs trading is most profitable for utility stocks.
	Do and Faff (2010)	US industry groups	1962 – 2009	Transaction price	Daily	Utility and financial stocks are the most profitable. Finer industry classification can increase profitability.
	Mori and Ziobrowski (2011)	US REIT (real estate investment trusts)	1987 – 2008	Transaction price	Daily	The REIT market is more profitable for pairs trading than the general market during the 1993 – 2000 period.
	Do and Faff (2012)	US industry groups	1963 – 2009	Transaction price	Daily	The best pairs earn 3.4% p.a. on average.
Strict form	Schultz and Shive (2010)	US dual-class shares	1993 – 2006	Transaction price and bid-ask quote	2-minute	Pairs trading between share classes of the same firm generates abnormal profits after transaction costs.
	Alsayed and McGroarty (2012)	UK stocks and ADRs	2011	Bid-ask quote	Tick	Pairs trading activities help maintain the stock-ADR price parity.
	Broussard and Vaihekoski (2012)	Finnish common and preferred stocks	1987 – 2008	Transaction price	Daily	Pairs trading between common and preferred stocks of the same company can create significant profits.
	Marshall et al. (2013)	US equity ETFs	2001 – 2010	Bid-ask quote	Tick	The average profit of pairs trading is 6.57% p.a. after considering the bid-ask spread.

3. Data

We focus on the most liquid pair of US gold ETFs, namely SPDR Gold Shares (ticker symbol GLD, provided by State Street Global Advisors) and iShares Gold Trust (ticker symbol IAU, provided by BlackRock). We choose the most liquid instruments because of two reasons. Firstly, liquid securities are traded more often, which enables the timely capture of short-lived arbitrage opportunities. Secondly, liquidity is an important determinant of the profit potential since actively traded assets allow more money to be made. Introduced on 18th November 2004 and 21st January 2005 respectively, both GLD and IAU aim to track the price performance of gold bullion. We collect quote data (time-stamped to milliseconds) from February 2005 to May 2010 from Thomson Reuters Tick History. The dataset ends in 2010 because of data unavailability. We have access to a certain amount of tick data and we decide to focus on the early period since the introduction of the more recent ETF (IAU), which may be more interesting than the period when it has become more mature, and 2010 is as far as we can get. Similar to Marshall et al. (2013), we consider only the core trading session (i.e. 9:30am – 4pm) to maximise liquidity.

To address potential errors in the data, we follow the data cleaning process used by Schultz and Shive (2010) and Marshall et al. (2013). Letting the bid and ask subscripts denote the bid price and ask price respectively, an observation at time t is removed if at least one of the following conditions is met:

$$
\begin{aligned}
&GLD_{bid,t} \ \geq \ GLD_{ask,t} \quad \text{or} \quad IAU_{bid,t} \ \geq \ IAU_{ask,t} \\
&GLD_{bid,t} \ \leq \ 0.25 \times GLD_{ask,t} \quad \text{or} \quad IAU_{bid,t} \ \leq \ 0.25 \times IAU_{ask,t} \\
&\left| ln\left(\frac{GLD_{bid,t}}{GLD_{bid,t-1}}\right) \right| > \ 0.25 \quad \text{or} \quad \left| ln\left(\frac{GLD_{ask,t}}{GLD_{ask,t-1}}\right) \right| > \ 0.25 \quad \text{or} \\
&\left| ln\left(\frac{IAU_{bid,t}}{IAU_{bid,t-1}}\right) \right| > \ 0.25 \quad \text{or} \quad \left| ln\left(\frac{IAU_{ask,t}}{IAU_{ask,t-1}}\right) \right| > \ 0.25 \\
&\frac{GLD_{bid,t}}{IAU_{ask,t}} \ > \ 1.5 \quad \text{or} \quad \frac{IAU_{bid,t}}{GLD_{ask,t}} \ > \ 1.5
\end{aligned}
\tag{1}
$$

Following Marshall et al. (2013), we also remove quotes posted during the first and last five minutes of trading. However, unlike them, we do not exclude the flash crash (6th May 2010) because unexpected situations are an important part of trading. Initially, there are 144,836,859 observations. After cleaning our data, there remain 144,099,869 valid observations. Table 2 presents the descriptive statistics of the clean data. Ranging from -0.55% to 0.49%, the mid-quote log returns are negatively skew, leptokurtic and non-normal as shown by the significant Jarque-Bera statistic.

Table 2. Descriptive statistics of mid-quote returns. The returns are in percentage. *** superscript denotes significance at 1%.

	GLD	IAU
Mean	1.36E-05	3.91E-05
Median	0	0
Maximum	0.48	0.49
Minimum	-0.52	-0.55
Standard deviation	0.04	0.05
Skewness	-0.18	-0.29
Kurtosis	16.7	19.53
Jarque-Bera normality	4663840 ***	6126751 ***

4. Methodology

4.1. Pairs trading based on full convergence

We apply a trading strategy similar to that of Marshall et al. (2013) with some adjustments as follows.

1. At time t_0, we $$\begin{cases} \text{sell GLD and buy IAU} \quad \text{if} \quad \frac{GLD_{bid,t_0}}{IAU_{ask,t_0}} \geq 1.002 \\ \text{sell IAU and buy GLD} \quad \text{if} \quad \frac{IAU_{bid,t_0}}{GLD_{ask,t_0}} \geq 1.002 \\ \text{do nothing otherwise} \end{cases} \quad (2)$$
2. We use contingent marketable limit orders so that each ETF trade is executed only if the other ETF trade can be executed at a pre-determined price. Such execution ensures that the exact mispricing observed is captured. The trigger value of 1.002 helps exclude a large number of small mispricings.
3. Regarding execution speed, Hasbrouck and Saar (2013) find that high-frequency traders in the US stock market can execute their orders in two milliseconds and their sample period starts in October 2007. Moreover, the US Securities and Exchange Commission documents in 2010 the ability of traders to operate in the microsecond environment (Goldstein et al., 2014), which suggests that execution speed tends to increase over time. Therefore, we use the speed of two milliseconds in our sub-sample starting in October 2007, which is conservative towards the end of the sample. Regarding the period before October 2007, the introduction of the Hybrid Market trading system in October 2006 allows US equity traders to reduce their

execution time from 10 seconds to under one second (Hendershott and Moulton, 2011). As a result, we use the execution speed of 10 seconds for our sub-sample before October 2006 and one second for our sub-sample from October 2006 to September 2007. For comparison purposes, we also consider the execution time of 15 seconds used by Marshall et al. (2013). Our actual pairs trade is opened at the first quote set available after the entry signal plus the time required for execution. If the ETF prices have moved against us during execution, this trade will not be opened.

4. At time t_1, we $\begin{cases} \text{close 'short GLD} - \text{long IAU' trades} & \text{if } \frac{IAU_{bid,t_1}}{GLD_{ask,t_1}} \geq 1 \\ \text{close 'short IAU} - \text{long GLD' trades} & \text{if } \frac{GLD_{bid,t_1}}{IAU_{ask,t_1}} \geq 1 \end{cases}$ (3)
5. The trade is actually closed at the first quote set available after the exit signal plus the time required for execution. Unlike the trade entry, even if there has been adverse price movement during execution, our trade will still be closed because it is important to be able to exit the trade, even at the expense of lower profits.
6. Following Marshall et al. (2013), the trading process above employs only fresh quote sets in which quotes of both ETFs have changed in the last five minutes.
7. To ensure that our trades take place within the core trading session, from 3:50pm in any trading day (i.e. ten minutes before the core trading session ends), any existing pairs trade will be closed at the first available quote set. Because of this mandatory position liquidation, we do not enter new trades after 3:30pm to allocate sufficient time for each trade. Strict-form pairs trades are generally short (e.g. Alsayed and McGroarty, 2012, Marshall et al., 2013) so 20 minutes should suffice.
8. The one-way commission per share traded at \$1.00 or more (GLD and IAU have always been traded above \$1.00) is 0.1 – 0.3 cent, depending largely on trading activities of the trader and the order type used [4]. Given the activities of high-frequency traders and the order type used in our analysis, we use the commission of 0.2 cent.
9. Regarding the short-selling cost, Engelberg et al. (2008) find that it is not a major friction. Moreover, Stratmann and Welborn (2012) state that the average short-selling cost of more than a thousand ETFs including many illiquid ones is only

[4] The fee schedule is available at https://www.nyse.com/publicdocs/nyse/markets/nyse-arca/NYSE_Arca_Marketplace_Fees.pdf

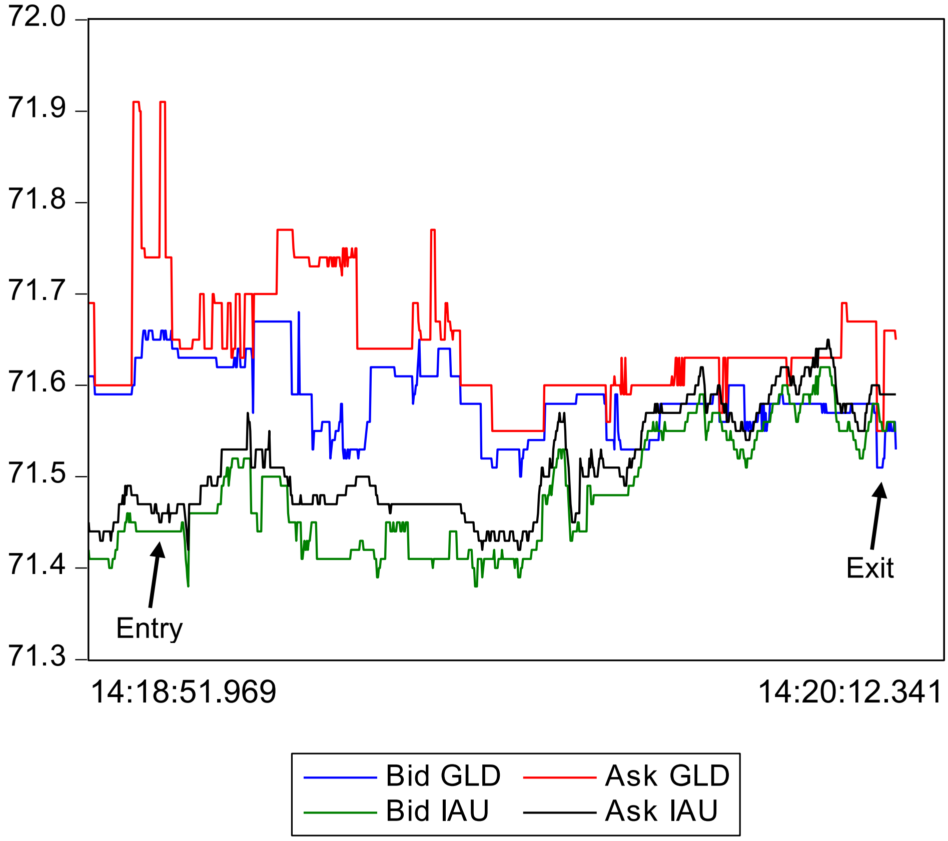
0.0035% per day. The cost of short-selling the highly liquid gold ETFs should be much lower and hence negligible.

10. Regarding short-selling restriction, a security cannot be short-sold at or below the current best bid once its price has declined by 10% or more from the close price of the previous day [5]. This restriction can only be triggered in the core trading session (not in the sessions before or after core trading) and will last until the end of the next trading day. We exclude all arbitrage opportunities which require short sales of GLD (IAU) while GLD (IAU) is under restriction.
11. Finally, each position requires a 50% margin (earning zero return) (Mitchell et al., 2002).

Figure 3 illustrates an example of our pairs trading. This trade takes place on 18th September 2007. At 14:18:51, when the two ETFs have shown sufficient divergence, the trade is opened by selling the overpriced GLD at bid price and buying the underpriced IAU at ask price. At 14:20:12, when they have converged (i.e. ask GLD has crossed bid IAU), the trade is closed at a profit by buying back GLD at ask price and selling IAU at bid price.

[5] The short-sales information is available at
http://www1.nyse.com/pdfs/8764_NYSEArca_FAQ_110225.pdf

Figure 3. Pairs trading example. The vertical axis shows the price in US dollar.



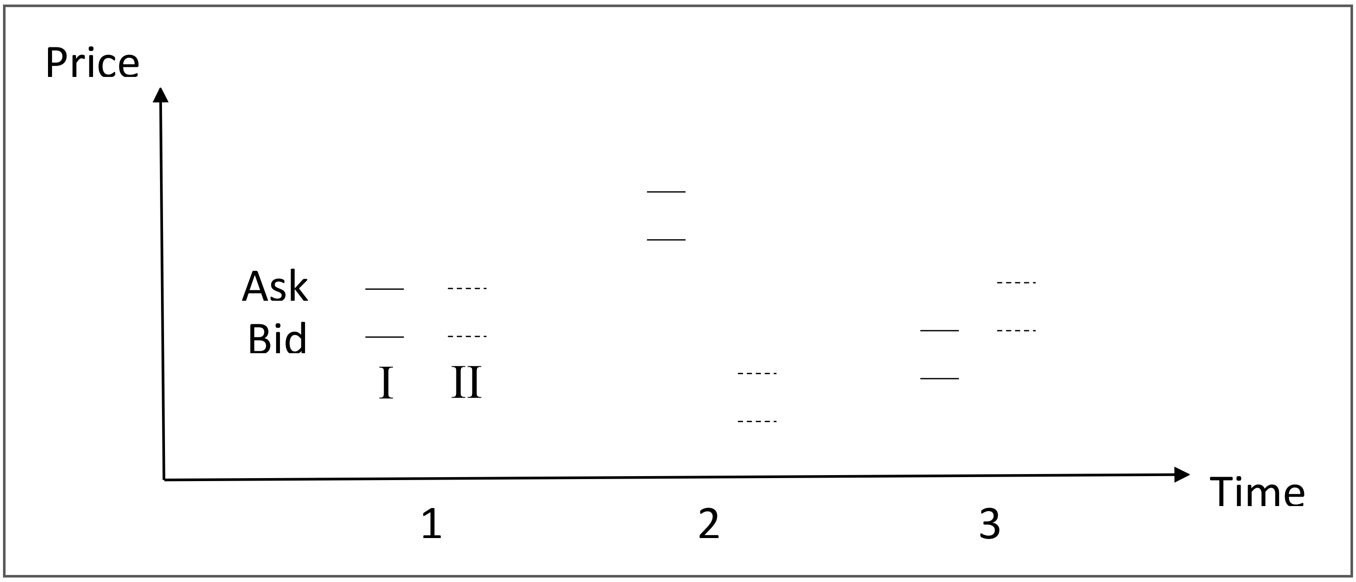
4.2. Pairs trading based on partial convergence

The standard pairs trading rule in the literature (e.g. Gatev et al., 2006, Do and Faff, 2010, Jacobs and Weber, 2015) requires complete elimination of the relative mispricing (i.e. trades will be closed only if the two stocks converge completely). We examine an alternative exit rule, namely partial convergence, which only requires partial elimination of mispricing. Specifically, trades are closed after a certain profit target has been reached during convergence.

4.2.1. Justification for partial convergence

There are ex-ante reasons why the partial-convergence strategy may be better than the full-convergence strategy. Firstly, we are motivated by findings in the literature regarding the speed of convergence during a pairs trade. Kondor (2009) finds that convergence during a pairs trade is increasingly slow because during convergence, the gap becomes smaller and less appealing to new arbitrageurs. Therefore, fewer traders will attempt to exploit the tightening gap and the gap takes longer to disappear due to the lack of trading pressure. Similarly, Alsayed and McGroarty (2012) show that arbitrageurs face uncertainty about the duration of trades and this uncertainty depends on the convergence target. Full convergence is not worth waiting for if the increased duration uncertainty outweighs the extra profit, which may explain why some mispricings are not exploited. As a result, the full-convergence strategy may not be optimal. Secondly, the partial-convergence strategy has inherent advantages compared to the full-convergence strategy, namely (i) we can capture more opportunities because a partial target allows trades to end quickly and thus allows the funds to become available quickly for the next trade and (ii) we can eliminate the possibility of the current winning trade turning into a losing trade while waiting for full convergence. Thirdly, the full-convergence exit rule employed in the literature may be unnecessarily strict, as illustrated in Figure 4.

Figure 4. A simplified pairs trade. The vertical axis shows the price and the horizontal axis shows the time which includes three periods. I and II are two ETFs tracking the same index or asset. There are two prices associated with each ETF, namely the bid and the ask.



In period 1, the two ETFs are in equilibrium and since they track the same index or asset, their prices should be similar. In period 2, they diverge temporarily, which triggers a pairs trade (i.e. selling I at the bid and buying II at the ask). In period 3, we close the trade only when they converge completely (i.e. the ask of I is equal to the bid of II) so that we can close both positions at the same price. However, the situation in period 3 is not the equilibrium in period 1. In fact, after returning to equilibrium, they must move further before we can exit the trade. Although the ETFs should return to equilibrium, there is no reason to expect them to move further after that. Therefore, the requirement of full convergence described above may be too strict and it is natural to relax it by using the partial-convergence exit rule.

4.2.2. The partial-convergence exit rule

The previous trading strategy in section 4.1, which is based on full convergence (evidenced by the exit condition in step 4), will be repeated with our partial-convergence exit condition as follows. Let us start with the 'short GLD – long IAU' trade. Letting P denote the profit ($) from convergence (excluding commission) of a given trade, α denote the profit target defined as a percentage of the profit from full convergence ($0 < \alpha < 1$), α and 100% subscripts denote the case of partial and full convergence, t_0 and t_1 denote the time of trade entry and exit; we have the following system.

$$\begin{aligned} P_\alpha &= GLD_{bid,t_0} - IAU_{ask,t_0} + IAU_{bid,t_1,\alpha} - GLD_{ask,t_1,\alpha} \\ P_{100\%} &= GLD_{bid,t_0} - IAU_{ask,t_0} + IAU_{bid,t_1,100\%} - GLD_{ask,t_1,100\%} \\ P_\alpha &= \alpha \times P_{100\%} \end{aligned} \tag{4}$$

It follows that

$$IAU_{bid,t_1,\alpha} - GLD_{ask,t_1,\alpha} = (\alpha - 1)\left(GLD_{bid,t_0} - IAU_{ask,t_0}\right) + \alpha\left(IAU_{bid,t_1,100\%} - GLD_{ask,t_1,100\%}\right) \tag{5}$$

By definition of full convergence, for 'short GLD – long IAU' trades, we have

$$IAU_{bid,t_1,100\%} \geq GLD_{ask,t_1,100\%} \tag{6}$$

It follows that

$$IAU_{bid,t_1,\alpha} - GLD_{ask,t_1,\alpha} \geq (\alpha - 1)\left(GLD_{bid,t_0} - IAU_{ask,t_0}\right) \tag{7}$$

This condition is our partial-convergence exit rule for 'short GLD – long IAU' trades. Similarly, at time t_1, we close 'short IAU – long GLD' trades if the following condition is met.

$$GLD_{bid,t_1,\alpha} - IAU_{ask,t_1,\alpha} \geq (\alpha - 1)\left(IAU_{bid,t_0} - GLD_{ask,t_0}\right) \tag{8}$$

While examining partial convergence, we mitigate the data mining issue by not testing too many values of α. Nevertheless, too few profit targets may provide less reliable results so we choose three evenly spaced targets (i.e. 25%, 50% and 75%). We also apply the Bonferroni adjustment to our tests. Moreover, data mining is less of an issue because we focus less on the optimal target but more on the general performance of partial convergence as a whole compared to full convergence.

4.3. Evaluation of pairs trading performance

Letting t_0 and t_1 denote the time of trade entry and exit, the profit (%) of a given pairs trade (whether based on full or partial convergence) is as follows.

$$
\begin{aligned}
&\text{short GLD} - \text{long IAU:} \quad \frac{GLD_{bid,t_0} - IAU_{ask,t_0} + IAU_{bid,t_1} - GLD_{ask,t_1} - 4\times 0.002}{0.5\times\left(GLD_{bid,t_0} + IAU_{ask,t_0}\right)}\times 100 \\
&\text{short IAU} - \text{long GLD:} \quad \frac{IAU_{bid,t_0} - GLD_{ask,t_0} + GLD_{bid,t_1} - IAU_{ask,t_1} - 4\times 0.002}{0.5\times\left(IAU_{bid,t_0} + GLD_{ask,t_0}\right)}\times 100
\end{aligned}
\tag{9}
$$

In each case, the last term in the numerator is the total commission which is four times one-way commission because a pairs trade requires four orders (two for entry and two for exit). The denominator is the margin requirement.

In terms of market efficiency, Jensen (1978) states that the most general interpretation of the Efficient Market Hypothesis is that if it is impossible to make economic profits (i.e. risk-adjusted returns after costs) using a given information set, the market is efficient regarding that information set. To arrive at implications for market efficiency, we calculate excess returns of pairs trading (over the risk-free rate of US dollar deposit) and adjust them for risks using the Sharpe (1994) and Sortino (2010) ratio. While the Sharpe ratio adjusts returns for general volatility, the Sortino ratio adjusts them for only downside volatility calculated from returns below the desired target return (DTR) which we set to zero. Because the Sortino ratio reflects the nature of risks more precisely than the Sharpe ratio, it might be the better risk adjustment. After considering risks using these ratios, if pairs trading outperforms the buy-and-hold strategy of these ETFs, the gold ETF market is inefficient. Letting $\overline{ER}$ denote the mean excess return, σ_{ER} denote the standard deviation of excess returns and T denote the number of observations; the Sharpe and Sortino ratios are as follows.

$$
Sharpe\ ratio = \frac{\overline{ER}}{\sigma_{ER}} = \frac{\frac{1}{T}\sum_{t=1}^{T} ER_t}{\sqrt{\frac{\sum_{t=1}^{T}\left(ER_t - \overline{ER}\right)^2}{T-1}}}
\tag{10}
$$

$$
Sortino\ ratio = \frac{\frac{\sum_{ER_t > DTR}(ER_t - DTR)}{T}}{\sqrt{\frac{\sum_{ER_t \le DTR}(DTR - ER_t)^2}{T}}} = \frac{\sum_{ER_t > DTR}(ER_t - DTR)}{\sqrt{T.\sum_{ER_t \le DTR}(DTR - ER_t)^2}}
\tag{11}
$$

To provide further insights into how the pairs trading returns relate to risks, we also apply another risk adjustment method to adjust the returns for exposure to risk factors. Following Engelberg et al. (2008), we regress the returns of pairs trading on (i) market-wide risks including Fama-French three-factor model (i.e. market return, size and value) plus two style factors (i.e. momentum and reversal) and (ii) liquidity risk whose proxy is the Treasury bill – Eurodollar (TED) spread (i.e. the difference between the US T-bill rate and the USD LIBOR rate). A high TED spread means that borrowing is difficult and thus liquidity is low and vice versa. If the market is efficient, the abnormal returns (i.e. intercept of the regression) should not be positive and statistically significant. Table 3 shows the summary statistics of these risk factors. The market return, size and momentum factors are negatively skew while the others are positively skew. All of the factors are leptokurtic and non-normal, as shown by the Jarque-Bera statistic.

Table 3. Descriptive statistics of the daily risk factors (in percentage). *** superscript denotes significance at 1%.

	Market	Size	Value	Momentum	Reversal	TED spread
Mean	0.006	0.01	0.011	-0.01	0.039	0.719
Median	0.08	0.01	0	0.07	-0.01	0.48
Maximum	11.35	3.85	4.8	7.05	11.22	4.58
Minimum	-8.95	-3.78	-4.22	-8.22	-7.16	0.09
Standard deviation	1.488	0.632	0.841	1.268	1.142	0.64
Skewness	-0.019	-0.06	0.623	-0.781	1.63	2.066
Kurtosis	12.176	8	11.248	10.629	23.127	8.921
Jarque-Bera normality	4704***	1398***	3888***	3388***	23228***	2913***

Engelberg et al. (2008) also mention divergence risk and default risk. However, in our ETF setting, divergence risk (i.e. the two stocks do not converge or they take a long time to do so) is low since (i) investors buy ETFs to track the underlying index or asset so the ETFs' management must try to minimise tracking errors to attract investors and (ii) ETF shares can be exchanged for the underlying asset or component stocks to benefit from potential mispricing, which should keep prices in equilibrium (Engle and Sarkar, 2006). On the other hand, default risk is also not an issue because ETFs are not company stocks and gold ETFs are backed by physical gold which can be redeemed by investors.

5. Results

Table 4 summarises the pairs trading performance of high-frequency traders, considering all trades and two subsets of winning and losing trades.

Table 4. Pairs trading performance with different profit targets. Our targets are defined as a percentage of full convergence (e.g. 50% means that trades are closed when the pair has converged by half of its initial divergence). Panel A shows the trading profit and the break-even transaction cost, panel B shows the risk adjustment and panel C shows the trade duration. The break-even transaction cost is the one-way commission per share which reduces the trading profit to zero. The excess return is equal to the trading profit plus the interest earned from the capital when not trading minus the risk-free rate. The Sharpe and Sortino ratio are compared between partial convergence and full convergence (vs. 100%) and between both strategies and the buy-and-hold benchmark (vs. b&h). We test for statistical significance of the Sharpe ratio following Pui-Lam and Wing-Keung (2008) and of the Sortino ratio using the bootstrapping technique in Levich and Thomas (1993). The speed of convergence is measured by the return per trading hour because the trading return relates directly to the distance of convergence. We report the break-even transaction cost, risk adjustment and speed of convergence based on all trades, which is more meaningful than based on only winners or losers. The * and *** superscript denote statistically significant difference at 10% and 1% level respectively.

	25%			50%			75%			100%		
	All trades	Winners	Losers	All trades	Winners	Losers	All trades	Winners	Losers	All trades	Winners	Losers
Number of trades	225	205	20	148	134	14	112	101	11	100	79	21
Panel A: Trading Profit (%)												
Total	14.760	17.559	-2.799	17.033	19.451	-2.418	16.525	18.926	-2.400	11.958	14.388	-2.430
Mean	0.066	0.086	-0.140	0.115	0.145	-0.173	0.139	0.187	-0.133	0.120	0.182	-0.116
Median	0.049	0.051	-0.060	0.102	0.105	-0.091	0.149	0.154	-0.049	0.090	0.194	-0.043
Standard deviation	0.237	0.228	0.234	0.285	0.276	0.197	0.323	0.318	0.187	0.209	0.169	0.177
Break-even TC (cents)	1.493	-	-	2.459	-	-	2.933	-	-	2.670	-	-
Panel B: Risk Adjustment												
Excess return (%)	14.005	-	-	16.279	-	-	15.770	-	-	11.204	-	-
Sharpe ratio (vs. 100%)	0.092	-	-	0.105	-	-	0.103	-	-	0.115	-	-
Sortino ratio (vs. 100%)	0.498 *	-	-	0.616 ***	-	-	0.648 ***	-	-	0.454	-	-
Sharpe ratio (vs. b&h)	0.092	-	-	0.105	-	-	0.103	-	-	0.115 *	-	-
Sortino ratio (vs. b&h)	0.498 ***	-	-	0.616 ***	-	-	0.648 ***	-	-	0.454 ***	-	-
Panel C: Duration (hours)												
Total	49.618	26.291	23.327	109.557	69.603	39.954	191.143	122.103	69.040	253.252	169.096	84.156
Mean	0.221	0.128	1.166	0.740	0.519	2.854	1.606	1.209	3.836	2.533	2.140	4.007
Median	0.014	0.012	0.200	0.042	0.024	3.534	0.328	0.073	4.032	1.595	1.165	4.433
Standard deviation	0.668	0.343	1.724	1.444	1.178	2.020	2.176	1.950	2.070	2.413	2.366	2.026
Speed of convergence	0.297	-	-	0.155	-	-	0.086	-	-	0.047	-	-

Regarding full convergence (i.e. the 100% case), our number of trades per year is comparable to that of equity ETFs in Marshall et al. (2013). Able to execute their orders with minimal delay, high-frequency traders make a positive profit on 79% of their trades and generate an excess return of 11.2% over the sample period. For comparison purposes, we also consider pairs trading with the longer execution time of 15 seconds used by Marshall et al. (2013). The results in Table 5 show that faster execution can increase (i) the number of trades (because fast execution can often avoid adverse price movements during execution which require skipping trades according to the strategy), (ii) the percentage of winning trades, (iii) the total profit (i.e. we only break even with the slower execution) and (iv) the average profit per trade; while decreasing the average trade duration.

Table 5. Pairs trading performance based on full convergence with fast execution compared to slow execution.

	Slow execution			Fast execution		
	All trades	Winners	Losers	All trades	Winners	Losers
Number of trades	33	18	15	100	79	21
Panel A: Profit (%)						
Total	0.120	2.505	-2.384	11.958	14.388	-2.430
Mean	0.004	0.139	-0.159	0.120	0.182	-0.116
Median	0.002	0.077	-0.112	0.090	0.194	-0.043
Standard deviation	0.233	0.194	0.164	0.209	0.169	0.177
Panel B: Duration (hours)						
Total	107.802	67.530	40.271	253.252	169.096	84.156
Mean	3.267	3.752	2.685	2.533	2.140	4.007
Median	3.671	5.365	1.654	1.595	1.165	4.433
Standard deviation	2.435	2.688	2.028	2.413	2.366	2.026

On the other hand, Table 4 shows that when the profit target decreases from full convergence to partial convergence, the average trade duration decreases (because normally it takes less time to reach a closer target) so the number of trades increases (because if a trade ends quickly, the capital becomes available quickly and we can enter the next trade soon). Moreover, full convergence generally takes longer than partial convergence not only because of the distance of the target but also because convergence becomes more and more difficult as the two ETFs approach full convergence. The reason is that (i) on one hand, full convergence requires the bid price of one ETF to converge to the ask price of the other but (ii) on the other hand, since these ETFs track the same asset, they should have similar bid and ask price so the bid price of one should remain lower and not converge to the ask price of the other. This explanation is supported not only by findings in the literature that convergence during a pairs trade is increasingly slow (e.g. Kondor, 2009, Alsayed and McGroarty, 2012) but also by our measure of convergence speed. The speed of convergence is measured by the return per trading hour (i.e. total return divided by total duration of trades) since the total return (i.e. the numerator) relates directly to the distance of convergence and the total duration of trades (i.e. the denominator) is the time of convergence. We find that the partial-convergence strategy has higher speed of convergence than the full-convergence strategy. The difficulty in achieving full convergence may also explain why the winning rates of the partial convergence targets are higher than that of full convergence because a currently winning trade may turn into a losing trade while waiting for full convergence.

Regarding profitability, the total profit shows that partial convergence outperforms full convergence as a trade exit criterion, which means that it may be a good idea to give up some profits per trade in exchange for faster trades and the ability to enter more trades. The most profitable criterion is 50% and the least profitable one is 100%, generating an excess return of 16.28% and 11.2% respectively during the sample period. Moreover, the relationship between the profit target and total profit is not monotonic; initial reduction of the target (i.e. from 100% to 75% to 50%) increases the profit but further reduction (i.e. from 50% to 25%) decreases it. This means when reducing the profit target, the higher trading frequency can compensate for the lower per-trade profit, but only to some extent. However, to mitigate the data mining issue, we do not focus on finding the optimal partial target but instead we are interested in the outperformance of partial convergence in general compared to full convergence. The break-even transaction cost ranges from 1.49 cents (25% target) to 2.93 cents (75% target) and is relatively high compared to the applicable cost on the stock exchange (i.e. less than 0.5 cent), which suggests that the pairs trading strategy is robust to transaction cost. With regards to trade duration, all of its statistics (i.e. total, mean, median and standard deviation) increase monotonically with the profit target. Finally, the speed of convergence, measured by the return per trading hour, shows how much return is generated per unit of time in a trade and thus shows the effectiveness of in-trade capital utilisation. The higher convergence speed of the partial-convergence strategy allows it to use the trading capital during trades more effectively than the full-convergence strategy.

As for the risk-adjusted performance, the Sharpe ratios show that there is no significant difference between partial and full convergence but in terms of the Sortino ratio, partial convergence outperforms full convergence significantly at 1% level. Moreover, because the Sortino ratio considers only downside volatility instead of general volatility like the Sharpe ratio, it might be a more accurate reflection of risks and thus a more appropriate adjustment for risks. More importantly, both the Sharpe and Sortino ratio show that our pairs trading outperforms the buy-and-hold strategy (whose Sharpe and Sortino ratios are 0.055 and 0.08 respectively) and this outperformance is statistically significant at 10% for the Sharpe ratio and 1% for the Sortino ratio. The low significance of the Sharpe ratio may be because this ratio also penalises the upside volatility (i.e. favourable returns). Because pairs trading outperforms the buy-and-hold strategy on a risk-adjusted basis, the gold ETF market may be inefficient.

Table 6. Exposure of pairs trading returns to daily risk factors. This table shows the regression results of the trading returns from different convergence targets (i.e. 25%, 50%, 75% and 100%) on the risk factors, namely (i) the market-wide risks including Fama-French three-factor model (i.e. market return, size and value) plus two style factors (i.e. momentum and reversal) and (ii) the liquidity risk whose proxy is the Treasury bill – Eurodollar (TED) spread. The numbers in brackets show the standard error of the estimated coefficients. The *, ** and *** superscript denote statistical significance at 10%, 5% and 1% level respectively.

	25%	50%	75%	100%
Intercept	0.00011***	0.00013***	0.00012***	9.09E-05***
	(3.17E-05)	(3.28E-05)	(3.34E-05)	(2.14E-05)
Market	4.7E-06	2.85E-05	2.23E-05	6.24E-06
	(2.72E-05)	(2.81E-05)	(2.86E-05)	(1.82E-05)
Size	-4E-05	-3.6E-05	-4.2E-05	2.12E-05
	(5.13E-05)	(5.3E-05)	(5.39E-05)	(3.43E-05)
Value	-8.3E-05	-9.1E-05*	-0.00011**	-9.8E-05***
	(5.08E-05)	(5.24E-05)	(5.33E-05)	(3.39E-05)
Momentum	-5E-05	-6.8E-05*	-7.6E-05**	-7.9E-05***
	(3.37E-05)	(3.48E-05)	(3.54E-05)	(2.26E-05)
Reversal	2.23E-05	4.07E-05	4.47E-05	4.12E-05**
	(2.99E-05)	(3.09E-05)	(3.14E-05)	(2E-05)
TED spread	0.00011**	0.00016***	0.00018***	0.00017***
	(4.96E-05)	(5.12E-05)	(5.21E-05)	(3.32E-05)
Adjusted R^2	0.00342	0.01201	0.0141	0.03288

Although some risk factors are statistically significant, they do not explain the variation in trading returns well, as shown by the low adjusted R^2. The abnormal returns (i.e. intercept) are positive and significant at 1% level, which suggests that the pairs trading profits exceed compensation for risks and the market may be inefficient. The partial-convergence strategy (i.e. 25%, 50% and 75% target) is less exposed to the risk factors than the full-convergence strategy (i.e. 100% target), evidenced by less significant coefficients of the factors and lower adjusted R^2.

Figure 5 shows the cumulative wealth of high-frequency traders using different profit targets to exit trades, starting with \$100. The lines exhibit similar upward movements and low volatility. The first and second half of the sample period roughly corresponds to the period before and after the global financial crisis respectively. The cumulative wealth increases gradually in both the pre- and post-crisis period, which suggests that the pairs trading strategy performs similarly in both periods.

Figure 5. Cumulative wealth from different profit targets. The starting point is \$100. The vertical axis shows the wealth in US dollar.

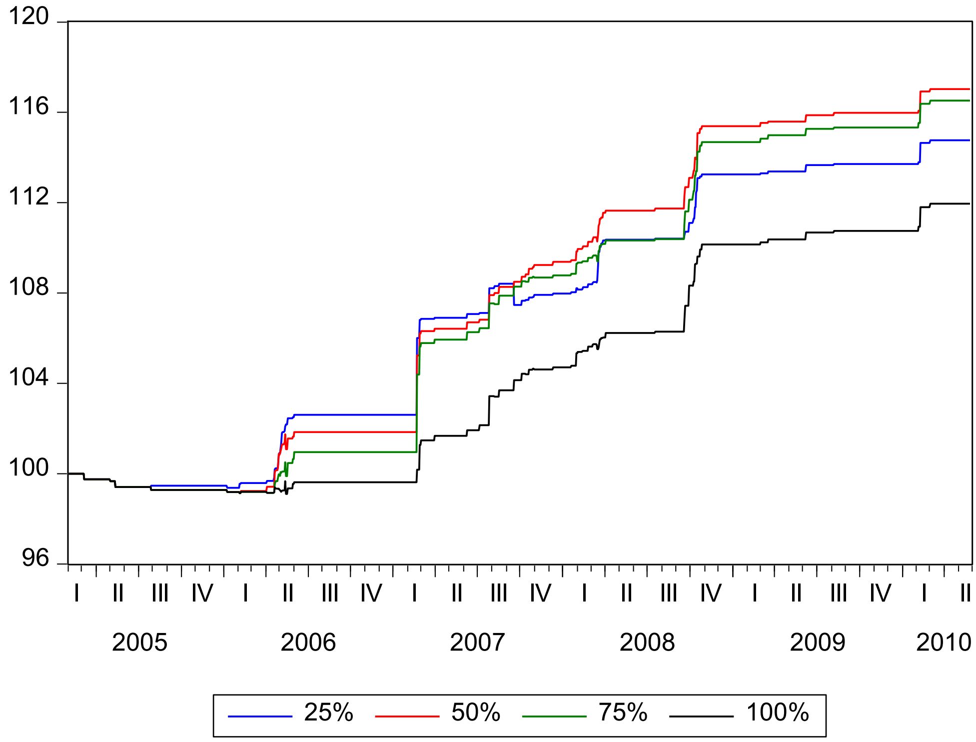


Figure 5 shows that the returns seem to cluster more in the middle than at the start or the end of the sample period. To examine this phenomenon, we divide the data into three sub-samples corresponding to the start, the middle and the end of the whole sample. Table 7 reports the test results for statistically significant differences in returns between the second sub-sample and the other two sub-samples. There are some significant differences between the first and second sub-sample and a possible explanation is that (i) pairs trading may be more profitable in bearish markets and (ii) the market is more bearish in the second sub-period than in the first one. Pairs trading may be more profitable in bearish markets because (i) it relies on the high correlation between two instruments to make them more likely to converge after their temporary price divergence and (ii) the correlation among financial assets increases in bearish markets (Longin and Solnik, 2001, Ang and Bekaert, 2002). We compare the return of the S&P 500 index (i.e. a proxy for the general market) in the first and second sub-period and find that the market is indeed much more bearish in the second sub-period (-6.64% return) than in the first one (15.91% return). However, the difference in returns between the second and the third sub-sample is not statistically significant.

Table 7. Variation of returns over time. This table shows the test results for statistically significant differences in returns between the second sub-sample and the other two sub-samples for both the partial-convergence and full-convergence strategy. The numbers are the test statistic. The *, ** and *** superscript denote statistical significance at 10%, 5% and 1% level respectively.

	First vs. second	Third vs. second
25%	-1.253	-0.777
50%	-1.958 *	-1.025
75%	-2.042 **	-0.684
100%	-3.121 ***	-0.256

Regarding the profit potential (i.e. how much money is available to obtain in the market), it depends on the liquidity of the two ETFs. During the sample period, the average daily trading volume and trading value of GLD are 3,339,335 shares and $290,291,641 respectively and those of IAU are 131,953 shares and $12,594,036 respectively. Both ETFs are highly liquid but because pairs trading requires trading both of them at the same time, the profit potential is determined by the less liquid instrument which is IAU. Even if only 1% of IAU's trading value (and an equivalent amount of GLD's trading value) is involved in pairs trading, the excess return of 2% p.a. generated by the baseline strategy can translate to an annual profit of more than $1.2 million. Moreover, this estimate is conservative because the trading volume is lower than the quoted volume which shows the total availability of these instruments. On the other hand, high frequency trading firms are unlikely to invest heavily in their infrastructure only to trade a single strategy so pairs trading can take advantage of the existing infrastructure used for other strategies and its upfront cost is reduced significantly. Finally, when the pairs trading strategy is established, more pairs can be added without incurring much extra cost.

Despite the potential issue of data mining, it is interesting to see whether there is an optimal convergence target. Figure 6 shows the performance of multiple targets from 5% increasing at 5% increments to 100%.

Figure 6. Performance of multiple targets. This figure shows the performance of multiple targets including the number of trades, the winning rate, the total profit and the total duration of trades. The whole sample period is also divided into two sub-samples of equal length for robustness check. The blue, red and green line represents the whole sample, the first half and the second half respectively. The horizontal axis shows the convergence targets (%).

Figure 6.1. Number of trades. The vertical axis shows the number of trades.

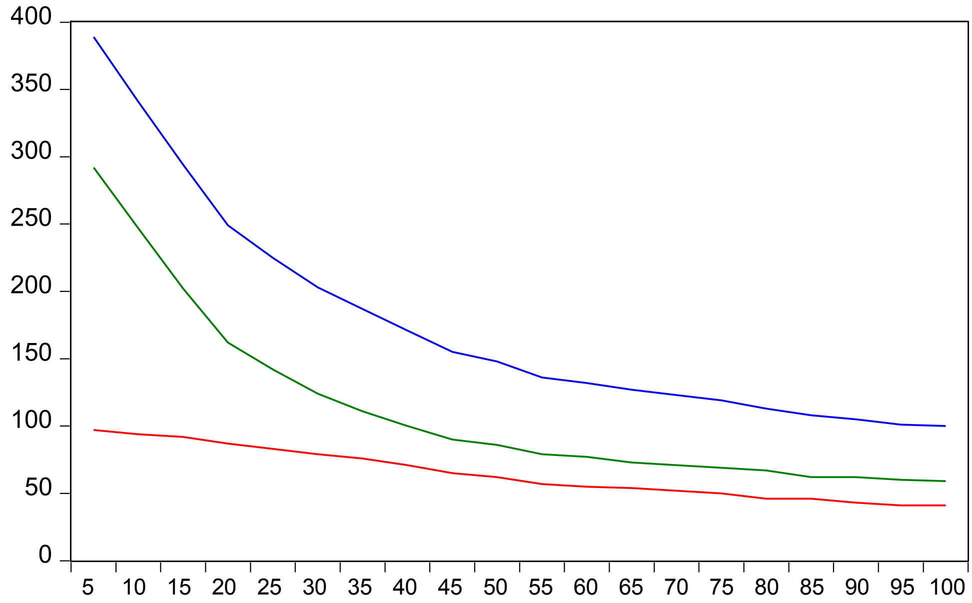


Figure 6.2. Winning rate. The vertical axis shows the winning rate (%). The winning rate is the number of profitable trades divided by the number of trades.

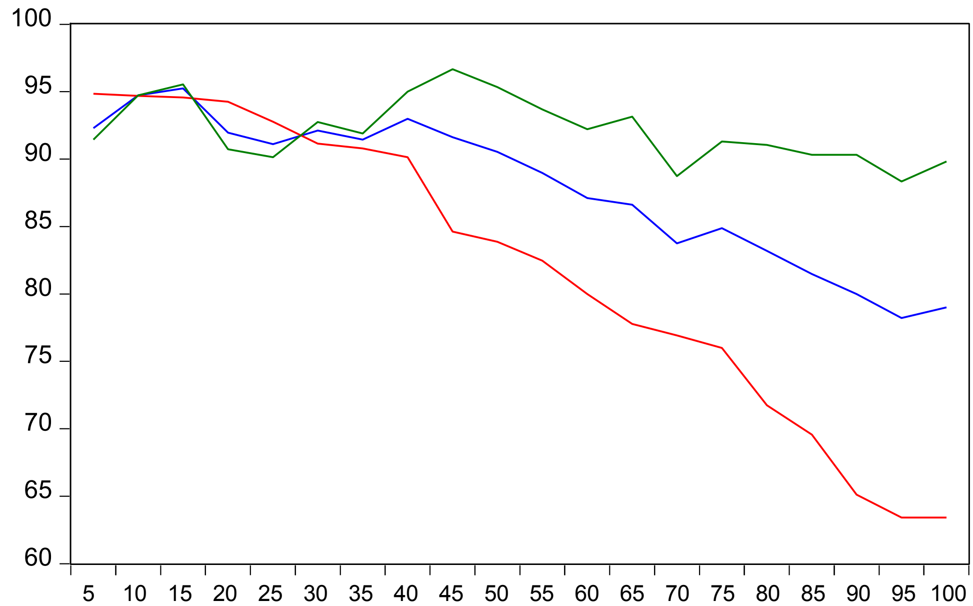


Figure 6.3. Total profit. The vertical axis shows the total profit (%).

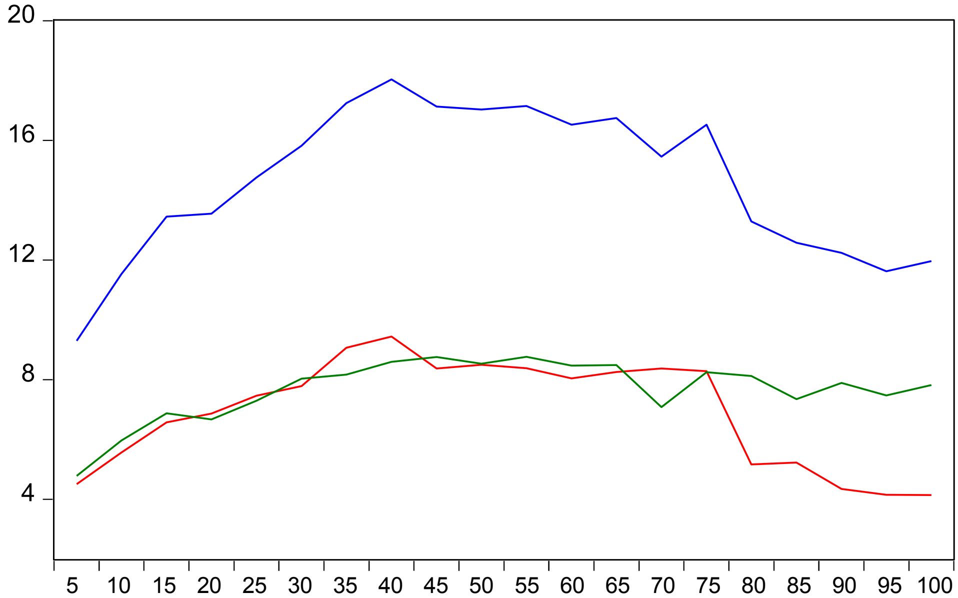
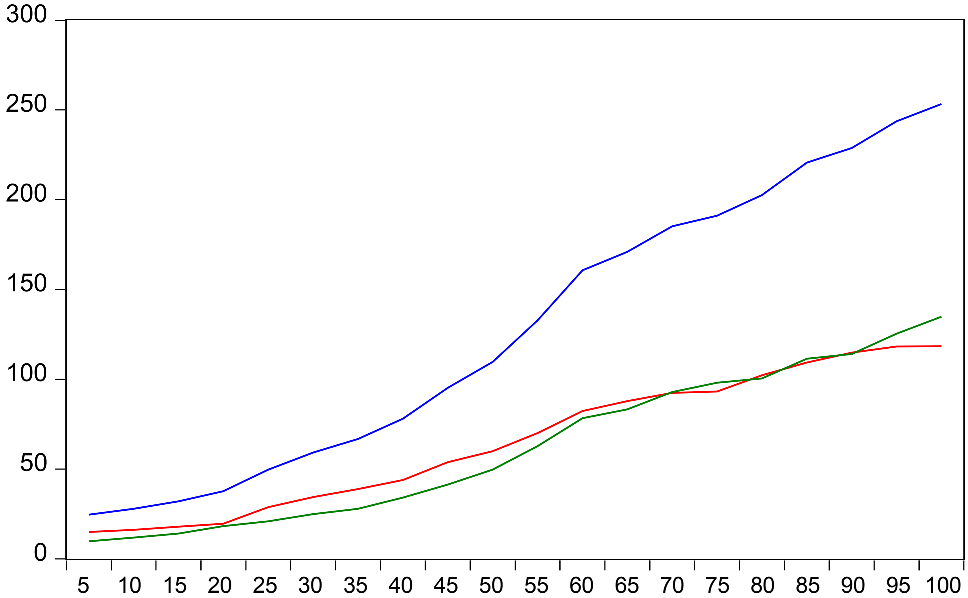


Figure 6.4. Total duration of trades. The vertical shows the total duration of trades (hours).



The patterns observed in the whole sample are also observed in the two sub-samples, which suggests that they are stable and robust. Both the number of trades and the winning rate tend to decrease when increasing the target, which shows that the partial targets can capture more opportunities than the full 100% target. Moreover, most partial targets lead to higher profits than the full target. The highest profit (i.e. 18% for the whole sample, 9.4% for the first half and 8.8% for the second half) is generated by the middle targets (i.e. 40% target for the whole sample and the first half and 45% target for the second half) as a combined result of a decent profit per trade and a relatively large number of trades. Finally, the duration of trades increases monotonically when increasing the target, showing that convergence is increasingly slow when using higher targets so partial convergence may be preferable to full convergence.

6. Conclusion

Using a large dataset of gold ETFs, we find that high-frequency traders, who can execute their orders with minimal delay, can profit from gold ETFs with pairs trading. To our knowledge, this paper is the first to analyse high-frequency pairs trading of gold ETFs. The fact that very fast order execution can capture more arbitrage opportunities and enhance profitability suggests that these opportunities are short-lived, which is consistent with other strict-form pairs trading studies (e.g. Alsayed and McGroarty, 2012, Marshall et al., 2013). According to Grossman and Stiglitz (1976) and Grossman and Stiglitz (1980), the profitability of our pairs trading may be compensation for the risks and costs involved in arbitrage and thus may motivate arbitrage activities. However, our excess return of 11.2% over the sample period or 2.1% p.a. is lower than that of equity ETFs in Marshall et al. (2013).

More importantly, we explain why the trade exit rule of full convergence used in previous studies may not be optimal and propose a rule based on partial convergence which outperforms the standard full-convergence rule. The trading return can be increased to 18.03% for the whole sample period or 3.38% p.a. by using our rule instead of the standard rule. Our partial convergence rule also enhances the risk-adjusted performance of pairs trading based on different risk adjustment methods (i.e. the Sharpe and Sortino ratio as well as regression analysis of the risk factors). In addition, the partial-convergence strategy has higher speed of convergence than the full-convergence strategy, which enables more effective use of the trading capital during trades. The outperformance of our rule is consistent in both the whole sample and the sub-samples with the optimal convergence target being around 40%, showing that pairs trading can exploit market inefficiency better when it requires only partial elimination of the relative mispricing. Finally, the pairs trading returns exceed compensation for risks, which suggests that the gold ETF market may be inefficient.

Acknowledgement

We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

References

ABREU, D. & BRUNNERMEIER, M. K. 2002. Synchronization risk and delayed arbitrage. *Journal of Financial Economics,* 66**,** 341-360.

ACKERT, L. F. & TIAN, Y. S. 2000. Arbitrage and Valuation in the Market for Standard and Poor's Depositary Receipts. *Financial Management,* 29**,** 71-87.

AKRAM, Q. F., RIME, D. & SARNO, L. 2009. Does the law of one price hold in international financial markets? Evidence from tick data. *Journal of Banking and Finance,* 33**,** 1741-1754.

ALSAYED, H. & MCGROARTY, F. 2012. Arbitrage and the Law of One Price in the market for American depository receipts. *Journal of International Financial Markets, Institutions and Money,* 22**,** 1258-1276.

ANG, A. & BEKAERT, G. 2002. International Asset Allocation With Regime Shifts. *Review of Financial Studies,* 15**,** 1137-1187.

BAUR, D. G. & LUCEY, B. M. 2010. Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold. *The Financial Review,* 45**,** 217-229.

BAUR, D. G. & MCDERMOTT, T. K. 2010. Is gold a safe haven? International evidence. *Journal of Banking and Finance,* 34**,** 1886-1898.

BLOSE, L. E. & GONDHALEKAR, V. 2013. Weekend gold returns in bull and bear markets. *Accounting & Finance,* 53**,** 609-622.

BREDIN, D., CONLON, T. & POTÌ, V. 2015. Does gold glitter in the long-run? Gold as a hedge and safe haven across time and investment horizon. *International Review of Financial Analysis,* 41**,** 320-328.

BROUSSARD, J. P. & VAIHEKOSKI, M. 2012. Profitability of pairs trading strategy in an illiquid market with multiple share classes. *Journal of International Financial Markets, Institutions & Money,* 22**,** 1188-1201.

CAGINALP, G., DESANTIS, M. & SAYRAK, A. 2014. The nonlinear price dynamics of U.S. equity ETFs. *Journal of Econometrics,* 183**,** 193-201.

CHEN, Z. & KNEZ, P. J. 1995. Measurement of Market Integration and Arbitrage. *Review of Financial Studies,* 8**,** 287-325.

CONRAD, J. & KAUL, G. 1989. Mean Reversion in Short-Horizon Expected Returns. *Review of Financial Studies,* 2**,** 225-240.

D'AVOLIO, G. 2002. The market for borrowing stock. *Journal of Financial Economics,* 66**,** 271-306.

DE LONG, J. B., SHLEIFER, A., SUMMERS, L. H. & WALDMANN, R. J. 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy,* 98**,** 703-738.

DO, B. & FAFF, R. 2010. Does simple pairs trading still work? *Financial Analysts Journal,* 66**,** 83-95.

DO, B. & FAFF, R. 2012. Are pairs trading profits robust to trading costs? *Journal of Financial Research,* 35**,** 261-287.

ENGELBERG, J., GAO, P. & JAGANNATHAN, R. 2008. An anatomy of pairs trading: the role of idiosyncratic news, common information and liquidity.

ENGLE, R. & SARKAR, D. 2006. Premiums-Discounts and Exchange Traded Funds. *Journal of Derivatives,* 13**,** 27-45.

FAMA, E. F. & FRENCH, K. R. 1997. Industry costs of equity. *Journal of Financial Economics,* 43**,** 153-193.

FROOT, K. A. & DABORA, E. M. 1999. How are stock prices affected by the location of trade? *Journal of Financial Economics,* 53**,** 189-216.

GAGNON, L. & KAROLYI, G. A. 2010. Multi-market Trading and Arbitrage. *Journal of Financial Economics,* 97**,** 53-80.

GATEV, E., GOETZMANN, W. N. & ROUWENHORST, K. G. 2006. Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies,* 19**,** 797-827.

GOLDSTEIN, M. A., KUMAR, P. & GRAVES, F. C. 2014. Computerized and High-Frequency Trading. *Financial Review,* 49**,** 177-202.

GROSSMAN, S. J. & STIGLITZ, J. E. 1976. Information and Competitive Price Systems. *American Economic Review,* 66**,** 246-253.

GROSSMAN, S. J. & STIGLITZ, J. E. 1980. On the Impossibility of Informationally Efficient Markets. *American Economic Review,* 70**,** 393-408.

GROSSMANN, A., OZUNA, T. & SIMPSON, M. W. 2007. ADR mispricing: Do costly arbitrage and consumer sentiment explain the price deviation? *Journal of International Financial Markets, Institutions & Money,* 17**,** 361-371.

HASBROUCK, J. & SAAR, G. 2013. Low-latency trading. *Journal of Financial Markets,* 16**,** 646-679.

HAUPTFLEISCH, M., PUTNIŅŠ, T. J. & LUCEY, B. 2016. Who Sets the Price of Gold? London or New York. *Journal of Futures Markets,* 36**,** 564-586.

JACOBS, H. & WEBER, M. 2015. On the determinants of pairs trading profitability. *Journal of Financial Markets,* 23**,** 75-97.

JEGADEESH, N. & TITMAN, S. 1995. Overreaction, Delayed Reaction, and Contrarian Profits. *Review of Financial Studies,* 8**,** 973-993.

JENSEN, M. C. 1978. Some anomalous evidence regarding market efficiency. *Journal of Financial Economics,* 6**,** 95-101.

KEARNEY, F., CUMMINS, M. & MURPHY, F. 2014. Outperformance in Exchange-Traded Fund Pricing Deviations: Generalized Control of Data Snooping Bias. *Journal of Financial Markets,* 19**,** 86-109.

KONDOR, P. 2009. Risk in Dynamic Arbitrage: The Price Effects of Convergence Trading. *Journal of Finance,* 64**,** 631-655.

LEVICH, R. M. & THOMAS, L. R., III 1993. The Significance of Technical Trading-Rule Profits in the Foreign Exchange Market: A Bootstrap Approach. *Journal of International Money and Finance,* 12**,** 451-474.

LONGIN, F. & SOLNIK, B. 2001. Extreme Correlation of International Equity Markets. *Journal of Finance,* 56**,** 649-676.

MARSHALL, B., NGUYEN, N. H. & VISALTANACHOTI, N. 2013. ETF arbitrage: Intraday evidence. *Journal of Banking & Finance,* 37**,** 3486-3498.

MITCHELL, M., PULVINO, T. & STAFFORD, E. 2002. Limited Arbitrage in Equity Markets. *Journal of Finance,* 57**,** 551-584.

MORI, M. & ZIOBROWSKI, A. J. 2011. Performance of Pairs Trading Strategy in the U.S. REIT Market. *Real Estate Economics,* 39**,** 409-428.

O'CONNOR, F. A., LUCEY, B. M., BATTEN, J. A. & BAUR, D. G. 2015. The financial economics of gold — A survey. *International Review of Financial Analysis,* 41**,** 186-205.

PUI-LAM, L. & WING-KEUNG, W. 2008. On testing the equality of multiple Sharpe ratios, with application on the evaluation of iShares. *Journal of Risk,* 10**,** 15-30.

PULLEN, T., BENSON, K. & FAFF, R. 2014. A Comparative Analysis of the Investment Characteristics of Alternative Gold Assets. *Abacus,* 50**,** 76-92.

SCHULTZ, P. & SHIVE, S. 2010. Mispricing of dual-class shares: Profit opportunities, arbitrage, and trading. *Journal of Financial Economics,* 98**,** 524-549.

SHARPE, W. F. 1994. The Sharpe ratio. *Journal of Portfolio Management,* 21**,** 49-58.

SHIN, S. & SOYDEMIR, G. 2010. Exchange-traded funds, persistence in tracking errors and information dissemination. *Journal of Multinational Financial Management,* 20**,** 214-234.

SHLEIFER, A. & VISHNY, R. W. 1997. The Limits of Arbitrage. *Journal of Finance,* 52**,** 35-55.

SORTINO, F. 2010. *The Sortino Framework for Constructing Portfolios,* United States of America, Elsevier.

STRATMANN, T. & WELBORN, J. W. 2012. Exchange-traded funds, fails-to-deliver, and market volatility.