ETF Arbitrage: Intraday Evidence

Ben R. Marshall
Massey University
B.Marshall@Massey.ac.nz

Nhut H. Nguyen*
University of Auckland
n.nguyen@auckland.ac.nz

Nuttawat Visaltanachoti
Massey University
N.Visaltanachoti@Massey.ac.nz

Abstract

We use two extremely liquid S&P 500 ETFs to analyze the prevailing trading conditions when mispricing allowing arbitrage opportunities is created. While these ETFs are not perfect substitutes, we show that their minor differences are not responsible for the mispricing. Spreads increase just before arbitrage opportunities, consistent with a decrease in liquidity. Order imbalance increases as markets become more one-sided and spread changes become more volatile which suggests an increase in liquidity risk. The price deviations are economically significant (mean profit of 6.6% p.a. net of spreads) and are followed by a tendency to quickly correct back towards parity.

JEL Classification: G1, G14
Keywords: Arbitrage, Pairs Trading, ETF

First Version: 23 September 2010
This Version: 24 January 2013

* Corresponding author: Nhut H. Nguyen, Department of Accounting and Finance, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand; telephone +64 9 3737599 ext. 83326; fax +64 9 3737406; e-mail n.nguyen@auckland.ac.nz. We thank Andrea Bennett, participants at the 2011 FMA Conference, and in particular, Dale WR Rosenthal (our session chair) and James Conover (our discussant), and participants at the 2011 New Zealand Finance Colloquium for valuable comments. All errors are our own.

ETF Arbitrage: Intraday Evidence

Abstract

We use two extremely liquid S&P 500 ETFs to analyze the prevailing trading conditions when mispricing allowing arbitrage opportunities is created. While these ETFs are not perfect substitutes, we show that their minor differences are not responsible for the mispricing. Spreads increase just before arbitrage opportunities, consistent with a decrease in liquidity. Order imbalance increases as markets become more one-sided and spread changes become more volatile which suggests an increase in liquidity risk. The price deviations are economically significant (mean profit of 6.6% p.a. net of spreads) and are followed by a tendency to quickly correct back towards parity.

JEL Classification: G1, G14
Keywords: Arbitrage, Pairs Trading, ETF

1. Introduction

Arbitrage is frequently described as one of the most important principles in finance. Numerous papers report situations of mispricing which appear to allow for arbitrage profits to be made (e.g. Froot and Dabora, 1999; Mitchell, Pulvino, and Stafford, 2002; and Gagnon and Karolyi, 2010).[1] Others point out that arbitrage is rarely, if ever, as risk-free and costless as the text-book definition suggests (e.g. De Long, Shleifer, Summers, and Waldman, 1990; Shleifer and Vishny, 1997). Much less is known about the intraday market characteristics that exist when arbitrage opportunities are created. It is this "microstructure of arbitrage" strand of the literature that we contribute to.

We consider mispricing between the SPDR Trust (ticker SPY) and iShare (ticker IVV).[2] The price deviations which result in the creation and removal of arbitrage opportunities are caused by purchases and sales of the ETFs in normal trading hours. This setting is ideal for an arbitrage microstructure study. Mispricing is easily identified and ETFs can be simply purchased and sold by all investors.[3] Other aspects of ETF arbitrage have been considered in the literature previously. Engle and Sarkar (2006) and Ackert and Tian (2010) report fluctuations in the mispricing or tracking error of ETFs relative to their underlying index. Engle and Sarkar (2006) note the creation / redemption feature of ETFs, which allows investors to exchange ETFs for stocks in the underlying index by submitting an order that is executed at the end of the day,

[1] The strategies used to capture this mispricing are variously referred to as "convergence trading", "pairs trading", and "statistical arbitrage." Papers applying these strategies include: Gatev, Goetzmann, and Rouwenhorst (2003), Engleberg, Gao, and Jagannathan (2010), Hogan, Jarrow, Teo, and Warachka (2004). Theoretical papers include Xiong (2001) and Bondarenko (2003).

[2] Vanguard listed an S&P 500 ETF in September 2010, but we do not include this in our analysis due to insufficient data. There are few other ETF pairs that track the same instrument. Where these do exist, the volume is typically heavily concentrated in one of the pairs, leaving relatively long intraday periods with no trading in the more illiquid security, which is problematic for a microstructure study such as this. An exception is the two ETFs on Gold.

[3] Dolvin (2010) shows the closing prices of the IVV and SPY sometimes deviate enough to allow arbitrage profits.

results in smaller deviations than in closed end funds which do not have this feature. Moreover, Richie, Daigler, and Gleason (2008) find the SPY can be used to exploit mispricing between the S&P 500 cash index and futures contracts.

Our evidence suggests that a fall in liquidity combined with an increase in liquidity risk contribute to the arbitrage opportunities. We show spreads increase prior to arbitrage opportunities, which is consistent with a decline in liquidity. Order imbalance increases, which is consistent with the market becoming more one-sided. The standard deviation of spread changes increases which is indicative of an increase in liquidity risk. Trade value also increases which, according to Johnson (2008), is evidence of an increase in liquidity volatility. Johnson (2008) explains the positive relation between liquidity risk and volume as follows (p. 411) "Intuitively, large changes in liquidity cannot occur without a lot of population flux, and a small amount of flux must imply a small change." Lastly, the standard deviation of trade-to-trade returns increases.

The S&P 500 ETFs are natural choices for an arbitrage microstructure study that considers the role liquidity plays in mispricing for a number of reasons. Firstly, they are highly liquid and widely followed. A large number of investors track these instruments and trade them (e.g. Elton. Gruber, Commer, and Li, 2002). The SPDR Trust (ticker SPY) is more liquid than any stock and the iShare (ticker IVV) is also highly liquid (above the 99th percentile of CRSP stocks based on value traded and below the 1st percentile based on spread). Any arbitrage opportunities can therefore be easily exploited. Secondly, it is well accepted that the risk to uninformed investors of trading against an investor with private or asymmetric information is considerably lower for investments in ETFs compared to investment in individual stocks.[4] This

[4] See Hamn (2011) footnote 1.

means that intraday changes in spreads are relatively (compared to stocks) more likely to be driven by liquidity changes than asymmetric information.

Thirdly, "convergence risk" is much lower than in most arbitrage settings. John Maynard Keynes once remarked "the market can stay irrational longer than you can stay solvent."[5] This creates a risk for arbitrageurs as simply identifying and trading on mispricing does not mean profits will be made quickly or at all.[6] The SPY and IVV compete for investor funds. Their ability to closely track the underlying index is an important aspect of this so the management of each fund have an incentive to minimize tracking error. Moreover, as Engle and Sarkar (2006) note, institutional investors can exchange each ETF for the underlying stocks. Larger ETF price divergence creates more incentive for this activity.

Another risk of arbitrage is fundamental risk. This refers to the fact that the two assets are not perfect substitutes or identical in all respects. It is extremely difficult to find assets that are completely identical. Dual class shares are likely to have different voting rights (e.g. Schultz and Shive, 2010). Dual listed stocks in different countries are frequently subject to different institutional features such as liquidity differences and index inclusion in one country (e.g. Froot and Dabora, 1999). Short-selling constraints in one country may play a role and there may also be "tax-induced investor heterogeneity" (e.g. Froot and Dabora, 1999, p. 215).

While minimized, fundamental risk does exist in our market setting. The SPY and IVV both have the aim of mirroring the S&P 500 index, but there may be small differences in the composition of the basket of securities each ETF uses to track the S&P 500. However, our results

[5] http://www.maynardkeynes.org/keynes-the-speculator.html
[6] Different aspects of this arbitrage risk have been documented. De Long, Shleifer, Summers, and Waldman (1990), focus on the risk of further price divergence due to the actions of irrational "noise" traders. Abreu and Brunnermeier (2002, p. 343) highlight that arbitrageurs still face synchronization risk or "uncertainty regarding the timing of the price correction" even when noise traders are not present. Mitchell, Pulvino, and Stafford (2002) refer to "horizon risk", and Shleifer and Vishny (1997) highlight the "margin risk" of leveraged positions having to be liquidated due to margin calls before the final convergence occurs.

suggest fundamental risk is not a major concern. We show NAV differences are generally smaller than the mispricing. There is no positive relation between the incidence of arbitrage opportunities and differences between the NAVs of the ETFs. There is also no statistically significant relation between the size of arbitrage profits and NAV differences. Moreover, the error correction results suggest that price deviations between the two ETFs result in convergence. When the price of each ETF deviates from that of the S&P 500 index or the other ETF the error correction mechanism pulls the ETF price back to the index level (other ETF price) in the next minute. Buyer-initiated trades are more common than seller-initiated trades in the underpriced ETF and seller-initiated trades are more prevalent than buyer-initiated trades in the overpriced ETF. This, together with the rapid removal of arbitrage profits (median time of 1-2 minutes), is further evidence that price deviations are seen as worth pursuing by investors.

We focus on mispricing of 0.2% and above (net of spreads) as we do not want the sample to be dominated by small mispricing that is offset by commissions and short sale costs. However, we also consider lower thresholds (0.1% and 0.15%). The mispricing we document is not a common occurrence. When the 0.2% threshold is used we find just 183 instances over the 2001 – 2010 period. Ninety (93) of these involve the SPY (IVV) trading at a higher price than the IVV (SPY). Median profits are 0.27% in both instances, while mean profits are 0.33% (0.32%) when the SPY (IVV) is initially trading at a premium. Mispricing is generally removed quickly. The median duration is 2.27 (0.92) minutes when the SPY (IVV) is trading at a premium. There are some outliers in the length of time it takes for arbitrage profits to be removed as some arbitrage opportunities occur late in the afternoon on a Friday.[7] This pushes the mean durations to 86.46 (3.92) minutes when the SPY (IVV) is trading at a premium. The arbitrage opportunities we document appear to be economically significant. Mean (median) p.a. profits, which we calculate

[7] We adopt the conservative assumption that positions are only closed in normal business hours.

by summing profits within a calendar year and then taking the mean (median) across years, are 6.57% (5.31%) after spreads. These compare favorably to the economic significance that has been attributed to other trading strategies. Most authors document profits prior to transaction costs. For instance, momentum profits from the Jegadeesh and Titman (1993) long-short strategy are often stated as being approximately 12% p.a., but this is prior to transaction costs. Lesmond, Schill, and Zhou (2004) note that much of this profitability comes from transactions in small illiquid stocks with relatively large transaction costs and that these costs subsume the momentum profits in many situations.

Schultz and Shive (2010) show mispricing of dual-class shares is typically driven by the more liquid share, and when this occurs trading volume in the underpriced (overpriced) share shifts from sales (purchases) to purchases (sales) and trading becomes more urgent. Schultz and Shive (2010) also show one-sided trades, rather than long-short trades, are used to eliminate the majority of the mispricing. We make several contributions beyond this important work. Firstly, we focus on highly liquid ETFs. Secondly, we consider a market setting with low convergence and fundamental risk – two factors which can inhibit arbitrageurs. Thirdly, we document the level of prevailing liquidity, liquidity risk and other security characteristics that exists in the minutes preceding, during, and following mispricing that is large enough for arbitrage opportunities.

The mispricing of ETFs has recently caught the attention of regulators. As the joint SEC / CFTC Flash Crash report[8] notes, ETFs accounted for 70% of all US-listed securities that declined by 60% or more during the May 6, 2010 Flash Crash. We hope this paper provides useful insight into the events that precede ETF mispricing.

[8] http://www.cftc.gov/stellent/groups/public/@otherif/documents/ifdocs/opa-jointreport-sec-051810.pdf

The remainder of this paper is organized as follows: Section 2 presents ETF characteristics, Section 3 contains a description of our data and the rules we use to identify divergent prices. Our results are presented in Section 4 and Section 5 concludes the paper.

2. SPY and IVV Characteristics, Co-movement, and Summary Statistics

As noted in the introduction, it is difficult to find two assets that are identical in all respects. The SPY and IVV are not identical. We outline different features in Section 2.1., provide summary statistics in Section 2.2., and give correlation and error correction results in Section 2.3. Movements in the IVV and SPY are highly correlated and there is a strong tendency for divergences in pricing to quickly error correct back to parity. We interpret these results as indicating investors view these assets as close substitutes. The Section 4 results, which show that changes in various ETF characteristics do not lead to the mispricing we consider occurring nor result in larger mispricing, reinforces this point.

2.1. ETF Characteristics

The SPY and IVV are both have the aim of mirroring the S&P 500 index, but there may be differences in the composition of the basket of securities each ETF uses to track the S&P 500. As noted in the SPDR prospectus, the trustees are permitted to allow a certain amount of "misweighting" between each stock in the SPDR and the S&P 500 index each day. It is only when the misweighting in a stock exceeds 150% of this amount that a re-balancing is required. There may also be differences in the actual securities held by each fund. The IVV fund documentation highlights it uses a "representative index sampling strategy". It does not always hold all 500 stocks of the S&P 500. [9]

Smith (2008) notes that the shares in the underlying securities that comprise the SPY are not permitted to be lent to third parties to earn revenue. However, components of the IVV can be lent. Smith (2008) also points out that dividends earned on the underlying securities are held in cash by the SPY managers and paid out quarterly. The IVV managers, in contrast, reinvest dividends received prior to the quarterly distributions. However, the SPY has a higher annual holdings turnover than the IVV.[10] The net expense ratio of the SPY ranged from 0.08% - 0.12% per annum over the 2001 – 2010 period we study, while the IVV expense ratio ranged from 0.09% - 0.10%. As shown in Figure 1, the maximum difference between these in any one year was 0.02%. Both the magnitude of and difference between these expense ratios are substantially less than the profits we document so their influence on mispricing is relatively minor.

[Insert Figure 1 Here]

2.2. ETF Co-movement

It is clear the S&P 500 ETFs diverge from the S&P 500 index on an intraday basis. Figure 1 shows deviations from parity (using one-minute intervals) between the IVV and S&P 500 index during the 1.30pm – 2.30pm period on September 29, 2008. This was a particularly volatile day of trading so the deviations are larger and more frequent than those on some other days, but this example does serve to illustrate that the ETFs frequently move from trading at a premium to a

[9] See the 2011 SPDR Prospectus https://www.spdrs.com/product/fund.seam?ticker=spy and the 2010 iShare Prospectus. http://us.ishares.com/product_info/fund/overview/IVV.htm.

[10] More information is available at http://finance.yahoo.com/q/pr?s=SPY and http://finance.yahoo.com/q/pr?s=IVV .

discount versus the S&P 500 index. The IVV (2010) prospectus states (p. 3): "The trading prices of the Fund's shares fluctuate continuously throughout trading hours based on market supply and demand rather than NAV. The trading prices of the Fund's shares may deviate significantly from NAV during periods of market volatility. While the ETF price deviations from the S&P 500 index depicted in Figure 2 are a necessary condition for arbitrage opportunities between the two ETFs to exist they are not sufficient. For this there needs to be deviations between the two ETFs that can be exploited. We outline the algorithm used to test for these arbitrage results in Section 3.

[Insert Figure 2 Here]

Table 1 Panel A shows that both movements in both ETFs are highly correlated with those of the S&P 500. The SPY – S&P 500 daily correlation is 0.98 while the IVV – S&P 500 correlation is 0.99. Engle and Granger (1987) state an error-correction representation exists if a set of variables are co-integrated. The unreported Johansen (1991) co-integration test shows that ETF prices and S&P500 at both the daily and one minute frequencies are co-integrated. We estimate the error correction model at both the daily and one minute frequencies as follows in equations 1a-1c.

$$\Delta SPY_t = \alpha_1 + \beta_1 \Delta S\&P500_{t-1} + \gamma_1 (SPY_{t-1} - \delta_1 S\&P500_{t-1}) + \varepsilon_{t1} \tag{1a}$$

$$\Delta IVV_t = \alpha_2 + \beta_2 \Delta S\&P500_{t-1} + \gamma_2 (IVV_{t-1} - \delta_2 S\&P500_{t-1}) + \varepsilon_{t2} \tag{1b}$$

$$\Delta SPY_t = \alpha_3 + \beta_3 \Delta IVV_{t-1} + \gamma_3 (SPY_{t-1} - \delta_3 IVV_{t-1}) + \varepsilon_{t3} \tag{1c}$$

Where $\gamma_1(SPY_{t-1} - \delta S\&P500_{t-1})$, $\gamma_2(IVV_{t-1} - \delta S\&P500_{t-1})$ and $\gamma_3(SPY_{t-1} - \delta IVV_{t-1})$ are the error correction terms that describes how the ETFs prices and S&P500 vary in short-run are consistent with a long-run co-integrating relationship; δs are co-integration parameters. The statistical significance of γ indicates that both ETFs error correct back to the S&P 500 at each of these frequencies.

The Panel B results show the daily SPY – IVV correlation is 0.99. Both ETFs error correct back to each other at the daily frequency and one minute frequency, although the error correction is much stronger in the IVV.

In Panel C we report summary statistics for ETF liquidity variables and NAV deviations. During the sample period the SPY had an average daily value traded of $13,374 million, while the IVV had an average daily value traded of $220 million. The SPY was more liquid than any NYSE / Amex / Nasdaq stock and the IVV was above the 99th percentile of stocks based on value traded. The average daily quoted spread of the SPY (IVV) is just 0.017% (0.046%), which makes the SPY the most liquid and places the IVV below the 1st percentile of all US stocks.

The liquidity (monthly value traded and spreads) of the SPY and IVV through time are plotted against the liquidity of the median CRSP stock in Figures 3a and 3b. These show that the SPY is consistently more liquid than the IVV, but both ETFs are much more liquid than the median stock in CRSP. The trading value of both ETFs is 100 times that of the median stock, while the spreads are typically one tenth of those of the median stock.

The mean daily NAV deviation (where a positive number indicates a higher SPY NAV) is -0.1%. The maximum is 0.36% and the minimum is -0.36%. In Section 4 we show that this level of NAV deviation is less than the mispricing we document. We also generate regression results which show that days with higher NAV deviations are not associated with either a greater likelihood of mispricing or larger mispricing.

[Insert Table 1 Here]

[Insert Figures 3a and 3b Here]

3. Data and Arbitrage Identification

We source high-frequency data from the Thomson Reuters Tick History (TRTH) database, which we access via the Securities Institute Research Centre of Asia-Pacific (SIRCA). TRTH data are used by academic researchers (e.g. Fong, Holden, and Trzcinka, 2011), central banks, hedge funds, investment banks, and regulators. More information on TRTH is available at their website.[11] We follow Hendershott, Jones, and Menkveld (2010) and limit our analysis to the post-decimalization period of February 2001 onwards. Our sample finishes in August 2010. We use data from the primary listing venue for each ETF. This was the AMEX prior to its takeover by NYSE Euronext in 2008 and NYSE Arca following that point. We source daily price, volume and spread data from CRSP to allow liquidity comparisons.

We adopt the data cleaning approach advocated by Schultz and Shive (2010), which involves deleting observations with any of the following characteristics:

1. Quotes posted outside normal trading hours.
2. Quotes posted in the first and last five minutes of regular trading.[12]
3. Bid price greater than or equal to the corresponding ask price for the same ETF.
4. Ask price at least four times larger than the bid price for the same ETF.
5. Bid-bid or ask-ask return for the same ETF is greater than 25% or less than -25%.
6. Bid price SPY / Ask price IVV > 1.5 or bid price IVV / ask price SPY > 1.5.

[11] http://thomsonreuters.com/products_services/financial/financial_products/quantitave_research_trading/tick_history

[12] This is broadly consistent with Schultz and Shive (2010) who omit quotes from the first and last four minutes of trading.

We also delete arbitrage observations that appear on May 6, 2010, the day of the Flash Crash, and any observations associated with dividend distributions of ETFs.[13]

Our assumption that trades only take place at the firm quotes in the market is conservative. As Schultz and Shive (2010) note, NYSE trades frequently occur at better prices. There are approximately 553 million SPY quotes and 220 million IVV quotes during the trading hours outlined above that pass our data cleaning process. This represents approximately 36,000 and 14,000 quotes per hour for the SPY and IVV respectively.

Our profit results are or are close to "net profits." The spread is accounted for and market impact costs are minimal for most traders with such highly liquid instruments. The only remaining execution costs appear to be commissions and short-selling costs, which are typically ignored in the pairs trading literature. Goldstein, Irvine, Kandel, and Wiener (2009) show the average one-way commission has declined over time to less than five cents per share in 2004. These have declined further since this point with 2010 NYSE documentation stating commissions of 0.15 – 0.3 cents per share for those with direct access to NYSE Arca.[14] Moreover, D'Avolio (2002) estimate S&P 500 stocks can be borrowed at less than 0.25% per annum, or approximately 0.0007% per day. Given these minimal costs, we suggest the profits we document are close to the actual profits large hedge funds / institutions could earn by exploiting these arbitrage opportunities. However, we apply a profit threshold of 0.2% as a conservative measure to ensure our sample is not dominated by the numerous smaller divergences.

The speed at which orders can be executed has increased over time. Bacidore, Ross, and Sofianos (2003) find the average order-to-execution time for NYSE orders in 1999 was 22.5

[13] ETFs typically distribute dividends, which they have collected from the stocks they hold, on a quarterly basis. These distributions rarely occur on the same day, which means that illusionary profit opportunities appear to be available in the period until both ETFs have gone ex-distribution.
[14] http://www.nyse.com/pdfs/NYSE_Arca_Marketplace_Fees_11_5_2010-Clean.pdf

seconds. Garvey and Wu (2009) document a mean (median) execution time of 12 (4) seconds for the 1999-2003 period. Hendershott and Moulton (2011) show execution times declined to less than one second in 2006 and Hasbrouck and Saar (2010) find some traders now react to market events within 2-3 milliseconds. However, we take the conservative approach and assume the orders take 15 seconds to be executed. For quotes to be included as valid quotes, both ETF quotes need to have been updated in the last five minutes. Our algorithm is as follows:

1. If bid SPY / ask IVV ≥ 1.002 or bid IVV / ask SPY ≥ 1.002 a *potential* arbitrage opening trade is identified. We assume a trader would submit contingent marketable limit orders upon observing the mispricing. These orders allow the trader to specify certain conditions before an order is executed.[15]
2. The *actual* arbitrage opening trade is opened at the first set of quotes for each ETF that appear 15 seconds after the *potential* opening arbitrage trade is identified. If there has been an adverse movement in the price of either ETF the orders will not be executed and the arbitrage trade will not be established.
3. In the case of the short SPY / long IVV trade, SPY ask and IVV bid prices are monitored. When a situation of IVV bid / SPY ask ≥ 1 is observed a *potential* arbitrage closing trade is identified. If a short IVV / long SPY trade was opened IVV ask and SPY bid prices are monitored and a *potential* arbitrage closing trade is identified when SPY bid / IVV ask ≥ 1. If 0.2% profit has been secured when the trade is opened we assume an arbitrageur is happy to close the trade at prices that do not result in further profit. We assume contingent marketable limit orders in a similar fashion to when the arbitrage was opened.

[15] Let's assume the SPY has a bid price of $155.08 and the IVV has an ask price of $154.57 (as per Figure 4a). This mispricing could be exploited by placing a contingent marketable limit buy order in the IVV at $154.57 that will only be executed if the SPY can be sold at $155.08 and a contingent marketable limit sell order in the SPY at $155.08 that will only be executed if the IVV can be purchased at $154.57. The interested reader should refer to: https://us.etrade.com/e/t/prospectestation/help?id=1301130000#b for more detail.

4. The *actual* arbitrage closing trade occurs at the first set of quotes for each ETF that appear 15 seconds after the *potential* closing arbitrage trade is identified.

We also test scenarios where an arbitrageur requires an opening profit of 0.3% and 0.4% and is happy taking a closing loss of up to 0.1% and 0.2% respectively to leave overall profits of 0.2% or more. These two strategies require less than total price convergence for arbitrages to be closed so are likely to be able to be closed more quickly.

Examples of arbitrage opportunities included in our results are provided in Figures 4a and 3b. As Figure 4a shows, at 15.38 on October 31, 2007, the bid price of the SPY was $155.08 and the ask price of the IVV was $154.57. These prices prevailed for 15 seconds so we assume an arbitrageur could sell at the SPY bid and buy at the IVV ask and lock in profit of 0.33% (155.08 – 154.57) / (0.5 × (155.08 + 154.57)), which is realized at convergence. The arbitrageur goes long the IVV and shorts the SPY. She maintains this position until 15.40 when the IVV bid and SPY ask converge at 154.73 for a period longer than 15 seconds. At this point, positions are closed at no further profit or loss, leaving overall profits of 0.33%. It is important to note this percentage profit calculation (and all others in this paper) is based on the assumption of Mitchell, Pulvino, and Stafford (2002), which is that a 50% margin (which earns no return) is required for both long and short positions. Actual profits would be larger if a lower margin is required and / or a return is earned on the margin that is posted. The example in Figure 3a is one of the 11.5% observations in our sample that occurs in the last hour of trading. Arbitrage opportunities are not concentrated in this period.

[Insert Figures 4a and 4b Here]

4. Results

Profit results are presented in Table 2. From Table 2 Panels A and B it is clear that it is almost equally likely for each ETF to be underpriced (overpriced). The SPY (IVV) is underpriced (overpriced) on 90 occasions and the IVV (SPY) is underpriced (overpriced) 93 times. The mean profits are also very similar (0.33% versus 0.32%). The maximum profit is 1.87% when the SPY (IVV) is underpriced (overpriced) and 1.27% when the IVV (SPY) is underpriced (overpriced).

The mean daily NAV deviation (see Table 1) is -0.1%. The maximum is 0.36% and the minimum is -0.36%. These statistics indicate that both mean profits and maximum profits are consolably larger than the mean and maximum NAV deviations, which suggests that the ETF mispricing is not solely determined by NAV deviations. We explore the NAV deviations and mispricing link further in Table 3.

We also test a number of variations of the base strategy and report these results in Appendix 1. These first two strategies require profits of 0.3% and 0.4% respectively for trades to be opened and convergence equating to losses of no more than -0.1% and -0.2% respectively for trades to be closed. These strategies therefore identify arbitrage opportunities with overall profits of 0.2% and larger. The results for these strategies are not mutually exclusive. For instance, an arbitrage opportunity has a profit 0.5% will appear in the base scenario and each of the variation strategies.[16] The results suggest profits are higher, on average, when the 0.3% and 0.4% opening thresholds are applied, but this higher limit results in fewer opportunities. For example, the mean profits of the 0.3% / -0.1% or 0.4% / -0.2% strategies when the SPY (IVV) is underpriced

[16] Each arbitrage is opened and closed within normal trading hours (excluding the first and last 5 minutes of trading) at quotes that prevail 15 seconds after each of the criteria are met. Our results are conservative as we only consider arbitrages generated with quotes posted in normal trading hours (excluding the first and last five minutes). In reality arbitrageurs could exploit divergent quotes after-hours and improve their profitability.

(overpriced) are 0.36% and 0.42% compared to 0.33% for the base strategy. We also consider strategies which lead to lower overall profit than 0.2%. Two alternatives are considered (0.10% and 0.15%). Reducing the profit threshold results in more instances of mispricing. There are 888 that generate profits of 0.10% or more and 374 instances of profit of 0.15% or more. Mean profits are lower than the 0.2% profit thresholds in Table 2 in each instance. Mean profits are 0.18% when the 0.1% threshold is applied and 0.25-0.26% when the 0.15% profit threshold is used.[17]

Panels C and D contain the average profit across all arbitrage opportunities by minute. Minute 0 refers to the profit on offer at the start of the minute when the arbitrage occurs. The means of 0.1% and 0.08% reflect the fact that the divergence required to open the arbitrage (0.2%) has not occurred yet. This occurs at various stages during minute 0. Minute 1 (-1) is the profit on offer at the start of the minute following (preceding) the minute where the 0.2% arbitrage threshold is reached. Statistically significant average arbitrage profits exist at the start of minute 1 but these have disappeared by the start of minute 2. The negative profits prior to time 0, which are statistically significant at times, do not indicate that positive profits could have been made from an opposite strategy of buying (selling) the overpriced (underpriced) ETF. They simply indicate that, on average, at that point in time the ask price of the underpriced ETF was higher than the bid price of the overpriced ETF.[18]

We present summary statistics for the length of time that divergent quotes allowing arbitrage profits persist for in Table 2 Panels A and B. Each duration is measured as the period of time divergent pricing lasts for plus 15 seconds (the delay that is imposed once the arbitrage

[17] The maximum profit differs from those in the 0.2% profit threshold because a lower profit threshold involves opening and closing arbitrage positions at different points. For instance, an arbitrage that is opened at 0.1% divergence might be followed by mispricing that leads to a 0.2% divergence. In this instance, with the benefit of hindsight, it would have been more profitable to delay opening the position.

[18] We remove a number of arbitrage opportunities from Table 2 Panels C and D and Table 4 due to their proximity to other mispricing. If these were not removed the pre and post period results could be affected by other opportunities.

opportunity is first identified). This ensures it is a conservative indication of the length of time an arbitrageur has to actually exploit the opportunity.

The divergent pricing which creates the arbitrage opportunities is removed relatively quickly. The median duration is 2.27 minutes when the SPY (IVV) is underpriced (overpriced) and 0.92 minutes when the IVV (SPY) is underpriced (overpriced). Arbitrages involving SPY (IVV) underpricing (overpricing) tend to last longer. The 75th percentile duration is 21.25 minutes when SPY is underpriced compared to just 1.64 minutes when IVV is underpriced. There are some extreme observations, due to arbitrage opportunities occurring at the end of the day (especially on Friday) which inflates the mean duration of profit opportunities. For instance, the maximum duration of 3,679 minutes was opened in the afternoon of Friday February 3, 2006 and closed on the morning of Monday February 6, 2006. However, even with these included the overall mean duration for the base strategy is 4 – 86 minutes.

To our knowledge, the rapid removal of arbitrage opportunities is quite different to the results of other studies on arbitrage in equity markets. For example, Froot and Dabora (1999) show mispricing in stocks listed in different locations can prevail for over four years[19], Mitchell, Pulvino, and Stafford (2002) show divergence in the price of a parent company and listed subsidiary can last for over five months[20], and Schultz and Shive (2010) show mispricing in class of stock with different voting rights can persist for two years.[21] However, the relatively quick restoration of efficient prices is consistent with Busse and Green (2002). They show prices converge to efficient levels following CNBC reports in one to fifteen minutes depending on whether the report is good or bad news. Moreover, Chorida, Roll, and Subrahmanyam (2005)

[19] See Froot and Dabora (1999) Figure 1 on page 193.
[20] See Mitchell, Pulvino, and Stafford (2002) Figure 1 on page 568.
[21] See Schultz and Shive (2009) Figure 2 on page 4.

find investors take between five and sixty minutes to restore prices to efficient levels following order imbalances.

Panel E contains summary statistics for the yearly profits earned from applying the ETF pairs trading strategy. Profits from individual transactions are cumulated within a calendar and summary statistics are calculated across the yearly totals. The mean (median) profits are 6.57% (5.31%).[22] These profits appear to be economically significant and compare favorably to those from trading strategies such as momentum. Momentum profits to the Jegadeesh and Titman (1993) momentum strategy are often reported as 12% p.a. However, this is the profit before bid-ask spreads have been accounted for, whereas the pairs trading profits we report are after spreads. Lesmond, Schill, and Zhou (2004) report that much of the momentum profits comes from transactions in small illiquid stocks so the profit drops considerably (and even disappears in many situations) when transaction costs are accounted for.

[Insert Table 2 Here]

The Appendix 3 results show there are fewer instances of mispricing in more recent times. In the first three full years in our sample (2002 – 2004) there are a total of 107 instances of mispricing, yet there are just 36 in the last three full years (2007 – 2009). The majority of this decline can be attributed to the subset of opportunities where the IVV (SPY) is relatively underpriced (overpriced). The number of these declined from 67 in the initial three-year period to six in the more recent three-year period. It is difficult to draw definitive conclusions regarding trends in the profits through time given the low number of observations in recent times, but there is some evidence of the few opportunities there were towards the end of the sample period

[22] This covers the period 2001 – 2009. There were no arbitrage opportunities in 2010.

generating larger profits. All else equal, the differential dividend treatment of the ETFs should have resulted in more SPY over pricing during the 2007 – 2009 period. Our results therefore suggest that other factors such as a decrease in liquidity and markets becoming more one-sided more than offset this effect.

In Table 3 we present results relating to the determinants of arbitrage mispricing and the magnitude of arbitrage profits. In Panel A we use a logit regression where a binary variable that equals 1 on days when an arbitrage opportunity is created and 0 otherwise is regressed on the ETF spreads, trading value, NAV differences and the VIX. In Panel B arbitrage profits are regressed on the same variables. Spread is the average of SPY and IVV quoted spreads throughout the trading day (multiplied by 10,000), Value is log of the average daily trading value across the two ETFs, NAV difference is the absolute value of the percentage difference in SPY and IVV NAVs. NAV data are from the ETF website and Morningstar and VIX data are from Global Financial Data.

The Panel A results show mispricing which is sufficient to generate arbitrage profits is more likely to occur on days when spreads are high. Moreover, there is a statistically significant negative relation between arbitrage opportunities and trading value in situations when the mispricing is driven by a relatively underpriced (overpriced) IVV (SPY). These results indicate that arbitrage opportunities are more likely to be generated on daily when markets are relatively less liquid. The coefficient of the NAV deviation variable is negative which suggests that arbitrage opportunities occur, on average, on days when NAV deviations are lower than average. Changes in NAV are not the driver of the arbitrage opportunities we document. Finally, VIX has a positive coefficient, which suggests arbitrage mispricing is more common on days when volatility is higher. The Panel B results indicate the size of the arbitrage profits is not related to either of the liquidity proxies, NAV deviations, or VIX. Summary statistics for the independent variables that are not presented in Table 1 are in Appendix 3.

[Insert Table 3 Here]

We now turn our attention to considering market characteristics immediately before each arbitrage opportunity is created. We calculate each variable at the time of the event. Time *t*-1 is the start of the minute prior to the arbitrage opportunity, *t*-2 is the previous minute, and so on. We measure each variable on the event day and at the same time of the day on the previous 20 trading days. The percentage change between the variable level on the event day and the mean on the previous 20 days is then calculated.[23] The median percentage change across all events is presented in Table 4. We have a pre-event period of *t*-5 to *t*-2, an event period of *t*-1 to *t*+1, and a post-event period of *t*+2 to *t*+5. Variables that are statistically significantly different to the previous period at the 10% level or better are in bold. For instance, if the *t*-1 to *t*+1 variable if different to the *t*-5 to *t*-2 variable the *t*-1 to *t*+1 number is in bold.

Spread is calculated for the first trade in each interval as two times the absolute value of the difference between the transaction price and the prevailing mid-price (e.g. Goyenko, Holden, and Trzcinka, 2009). Depth is the value of shares at the first level of the book (both the bid and ask) size at the start of each interval. Order Imbalance is calculated as the difference in the absolute value between buyer-initiated trades and seller-initiated trades divided by the sum of the two in each one-minute interval. Trades are classified as buyer- or seller-initiated using the Lee and Ready (1991) algorithm. The Return Standard Deviation is calculated based on the standard deviation of traded returns in each one-minute interval. Trade Value is the total value of trades

[23] Change is calculated for the sidedness as this is already expressed as a percentage.

during each minute interval, and Spread Standard Deviation is computed as the standard deviation of effective spreads in each interval**.**

The Table 4 results show that arbitrage opportunities occur at times when liquidity declines in both the SPY and IVV. Spreads are 41% - 51% higher in the event window (minute -1 to +1) than the equivalent time of day in the prior 20 trading days in the SPY. When the IVV (SPY) is underpriced (overpriced) spreads are 16%-18% larger. Moreover, three of the four spreads are statistically significantly larger in the [-1, +1] interval than they are in the [-5, -2] interval. Depth is statistically significantly lower in the event window than [-5, -2] interval and is 18% lower than the level from the previous 20 days in the IVV when the IVV is underpriced, but this decline in depth is not evident in other instances.

Order imbalance increases dramatically in the minutes surrounding an arbitrage opportunity. It is 17% - 41% higher in the [-1, +1] interval on event days than the previous 20 trading days and the event window order imbalance is higher than that of the previous interval in three out of the four instances. This is consistent with markets becoming more one-sided.

There is a surge in spread standard deviation in the IVV in the minutes surrounding arbitrage opportunities. These are 280% (300%) higher in the event window than they are at the same time in prior 20 trading days when the IVV is overpriced (underpriced). The spread standard deviation of the SPY also increases (24% - 61%) when arbitrage opportunities are created. Trading value in both the SPY and IVV also increases around arbitrage opportunities. Johnson (2008) points out that numerous empirical studies find volume and liquidity (e.g. spread) to be unrelated. He develops a frictionless model to explain this phenomenon. His model shows that liquidity reflects the risk bearing capacity or average willingness of traders to accommodate trades while volume represents the second moment of liquidity changes or liquidity risk. Pástor and Stambaugh (2003) and Acharya and Pederson (2004) highlight the importance of liquidity risk on asset pricing. Johnson (2008, p. 389) suggests that "liquidity risk is important from a policy perspective because of the danger posed by large drops in liquidity, which may lead to price distortions, disruptions in risk transfer, and possibly inefficient liquidation of real investments." Our results indicate the significant price distortion between the two ETFs occurs when liquidity dries up and liquidity risk increases, which is a period that the risk bearing capacity drops and such capacity is likely to fall further. We also find that the return standard deviation increases. Our findings are consistent with Deuskar and Johnson (2011) who show that unpredictable flow driven risk accounts for around half of market variance.

The results for the interval *t*+2 to *t*+5 indicate that spreads stay relative high in the minutes following each arbitrage opportunity, order imbalance, and return standard deviation declines in the SPY but not the IVV, and the spread standard deviation declines in the IVV but not the SPY.

[Insert Table 4 Here]

Divergent pricing in ETFs has recently captured the attention of regulators. The May 6, 2010 "Flash Crash" has been the catalyst for regulators to look at the reasons behind ETF pricing deviations from the indices they seek to mirror. In their joint report[24], the SEC and CFTC note ETFs made up 70% of all US-listed securities that plunged 60% or more on May 6. The SEC / CFTC report suggests one explanation for this might be institutions using ETFs as a quick way of reducing market exposure. Another possible explanation is the use of stop-loss market orders in ETFs by individual investors. Borkovec, Domowitz, Serbin, and Yegerman (2010, p. 1) suggest the failure in ETF price discovery during this event was due to "an extreme deterioration in

[24] The report is entitled "Preliminary Findings Regarding the Market Events of May 6, 2010" and is available here: http://www.cftc.gov/stellent/groups/public/@otherif/documents/ifdocs/opa-jointreport-sec-051810.pdf

liquidity, both in absolute terms and relative to individual securities in the baskets tracked by the funds."

We plot the deviations of the SPY and IVV from the level of the S&P 500 index at one-minute intervals during the 2.30pm – 3.30pm period on May 6, 2010 in Figure 5. Numbers greater (less) than zero indicate the ETF is trading at a premium (discount). The black bars represent the total SPY variation and the entire white bar (including the black portion) represents the IVV variation. Figure 5 makes it clear the IVV was more affected than the SPY. Its deviations from parity range from +4% to -10%. It is also evident the ETFs change from trading at a large premium one minute (2.45pm) to a large discount the next minute (2.46pm). On one occasion (2.49pm) the SPY is trading at a premium of over 2% and the IVV is trading at a discount of more than -2%. We do not include any observations from May 6, 2010 in our formal analysis.

[Insert Figure 5 Here]

In Table 5 we present results which provide insight into the convergence of prices following the arbitrage opportunities. We consider whether there is a difference in the number, volume, and dollar volume of buyer- and seller-initiated trades in the time between mispricing allowing arbitrage profits and their removal. We also look at aggressive buy (sell) trades, which we define as trades that go through at or above the prevailing ask price (at or below the prevailing bid price). The results indicate buyer-initiated trades are more likely in the underpriced ETF and seller-initiated trades are more common in the overpriced ETF. Moreover, aggressive buy trades are more prevalent in the underpriced ETF and aggressive sell trades are more likely in the overpriced ETF. The differences are statistically significant when the IVV is underpriced and the SPY is overpriced. Taken together, we suggest the Table 5 results provide evidence that arbitrageurs act to profit from the mispricing. Their actions lead to an ETF prices error correcting back to parity and short-lived mispricing.

[Insert Table 5 Here]

5. Conclusions

We document the intraday characteristics that prevail when mispricing which is large enough to create arbitrage profits. We show that liquidity declines (i.e. spreads increase) and order imbalance increases. The standard deviation of spread changes increase which suggests that liquidity risk increases. Trade value increases which, according to Johnson (2008), is also indicative of an increase in liquidity volatility. Return volatility is also higher when arbitrage opportunities occur.

Our results are based on two S&P 500 ETFs. The SPDR Trust (ticker SPY) is more liquid than any US stock based on both value traded and bid-ask spread and the iShare (ticker IVV) is in the 99^{th} (1^{st}) percentile of stocks based on value traded (spread). Having highly liquid instruments is important as it allows us to calculate meaningful statistics at high frequencies. The S&P 500 ETFs are natural instruments for other reasons. They are close substitutes for each other (low fundamental risk). The ETF providers have an incentive to minimize tracking error so as to attract more funds and holders can exchange the ETF for the underlying shares, which mitigates the risk that these ETFs will not converge following mispricing. While these ETFs are not perfect substitutes, we show that their minor differences are not responsible for the mispricing. We find that the ETFs have a daily return correlation of 0.99 and their deviations correct back following mispricing. Median durations are just 1-2 minutes. There is an increase in buyer-initiated trades (seller-initiated trades) in the underpriced (overpriced ETF) when mispricing occurs as arbitragers seek to profit from the mispricing. NAV differences between the ETFs do not drive the mispricing. Arbitrage opportunities are, however, more likely to be created when the market is more volatile. The arbitrage profits we document, which are net of bid-ask spreads, appear economically significant. They certainly compare favourably to the profits documents for the long-short momentum strategy.

The pricing of ETFs have recently captured the attention of regulators. During the May 6, 2010 Flash Crash a disproportionate number of ETFs were among those securities that experienced the largest price declines. The prices of ETFs also temporarily diverged by large amounts (greater than 10%) from their “true value” based on the indices they track. We hope that our research improves understanding of the factors that result in ETF mispricing on an intraday basis.

References

Abreu, Dilip and Markus K. Brunnermeier. (2002). Synchronization risk and delayed arbitrage. *Journal of Financial Economics*, 66, 341–360

Acharya, Viral .V. and Lasse Heje Pedersen. (2004). Liquidity risk and asset pricing. *Journal of Financial Economics,* 77, 375–410.

Ackert, Lucy F. and Yisong S. Tian. (2000). Arbitrage and valuation in the market for Standard and Poor's Depositary Receipts. *Financial Management*, 29(3), 71-87.

Bacidore, Jeffrey, Ross, Katharine., and George Sofianos. (2003). Quantifying market order execution quality at the NYSE. *Journal of Financial Markets*, 6, 281–307.

Bondarenko, Oleg. (2003). Statistical arbitrage and securities prices. *Review of Financial Studies*, 16(3), 875-919.

Borkovec, Milan., Domowitz, Ian., Serbin, Vitaly., and Henry Yegerman. (2010). Liquidity and price discovery in exchange traded funds: One of several possible lessons from the flash crash. ITG Investment Group. http://www.itg.com/news_events/ITG-Paper-LiquidityPriceDiscovery.pdf

Busse, Jeffrey and Clifford T. Green. (2002). Market efficiency in real time. *Journal of Financial Economics*, 65, 415-437.

D'Avolio, Gene. (2002). The market for borrowing stock. *Journal of Financial Economics*, 66 (2–3), 271–306.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J Waldmann. (1990). Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.

Deuskar, Prachi and Timothy Johnson (2011). Market liquidity and flow driven risk. *Review of Financial Studies* 24(3), 721-753.

Dolvin, Steven D. (2010). S&P ETFs: Arbitrage opportunities and market forecasting. *The Journal of Index Investing*, Summer, 107-116.

Elton, Edwin J., Gruber, Martin J., Comer, George, Kai Li. (2002). Spiders: Where are the bugs? *Journal of Business*, 75(3), 453-472.

Engle, Robert, F. and Clive W. J. Granger (1987). Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica* 55, 251-276.

Engle, Robert and Debojyoti Sarkar. (2006). Premiums-discounts and exchange traded funds. *Journal of Derivatives*, Summer, 27-45.

Engleberg, Joseph, Gao, Pengjie, and Ravi Jagannathan. (2010). *An anatomy of pairs trading: the role of idiosyncratic news common information and liquidity*. SSRN Working Paper: http://ssrn.com/abstract=1330689

Fong, Kingsley, Craig Holden, and Charles Trzcinka (2011). What are the best liquidity proxies for global research? SSRN working paper, http://ssrn.com/abstract=1558447.

Froot, Kenneth. A. and Emil. M. Dabora. (1999). How are stock prices affected by the location of trade? *Journal of Financial Economics*, 53, 189-216.

Garvey, Ryan and Fei Wu (2009) Intraday time and order execution quality dimensions. *Journal of Financial Markets*, 12, 203-228.

Gagnon, Louis and G. Andrew Karolyi. (2010). Multi-market trading and arbitrage. *Journal of Financial Economics*, 97, 53-80.

Gatev, Evan, Goetzmann, William N. and K. Geert Rouwenhorst. (2003). Pairs trading: Performance of a relative value arbitrage rule. *Review of Financial Studies*, 19(3), 797-827.

Goldstein, Michael A., Irvine, Paul., Kandel, Eugene., and Zvi Wiener. (2009). Brokerage commissions and institutional trading patterns. *Review of Financial Studies*, 22(12), 5175-5212.

Hamm, Sophia J.W., (2011). The effect of ETFs on stock liquidity. SSRN Working Paper: http://ssrn.com/abstract=1687914

Hasbrouck, Joel and Gideon Saar. (2010). Low-latency trading. SSRN Working Paper: http://ssrn.com/abstract=1695460

Hendershott, Terrence., Jones, Charles M., and Albert J. Menkveld. (2010). Does algorithmic trading improve liquidity? *Journal of Finance – Forthcoming*.

Hendershott, Terrence and Pamela C. Moulton. (2011). Automation, speed, and stock market quality: The NYSE's Hybrid. *Journal of Financial Markets*, 14, 568 -604.

Hogan, Steve., Jarrow, Robert., Teo, Melvyn., and Mitch Warachka. (2004). Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. *Journal of Financial Economics*, 73, 525-565.

IVV (2010). 2010 Prospectus to Shareholders. iShares S&P 500 Index Fund, August 2010.

Jegadeesh, Narasimhan, and Sheridan Titman. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48, 65-91.

Johansen, Søren. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6), 1551–1580.

Johnson, Timothy C. (2008). Volume, liquidity, and liquidity risk. *Journal of Financial Economics*, 87, 388-417.

Lee, Charles, and Mark Ready. (1991). Inferring trade direction from intraday data, *Journal of Finance*, 46,733–746.

Lesmond, David, Schill, Michael, and Chunsheng Zhou. (2002). The illusory nature of momentum profits. *Journal of Financial Economics*, 71, 349-380.

Mitchell, Mark., Pulvino, Todd., and Erik Stafford. (2002). Limited arbitrage in equity markets. *Journal of Finance*, 57(2), 551-584.

Pástor, Ľuboš, and Robert F. Stambaugh (2003). Liquidity risk and expected stock returns, *Journal of Political Economy*, 111, 642–685.

Richie, Nivine, Daigler, Robert T., and Kimberly C. Gleason. (2008). The limits to stock index arbitrage: Examining S&P 500 Futures and SPDRs. *Journal of Futures Markets*, 28(12), 1182-1205.

Schultz, Paul, and Sophie Shive. (2010). Mispricing of dual-class shares: Profit opportunities, arbitrage, and trading. *Journal of Financial Economics*, 98, 524-549.

Shleifer, Andrei. and Robert W. Vishny. (1997). The limits of arbitrage. *Journal of Finance*, 52, 35-55.

Smith, Casey. (2008). Investing in the S&P 500: SPY vs. IVV. Seeking Alpha. http://seekingalpha.com/article/109780-investing-in-the-s-p-500-spy-vs-ivv

SPY (2011). SPDR S&P 500 ETF Trust Prospectus, January 2011.

Xiong, Wei. (2001). Convergence trading with wealth effects: An amplification mechanism in financial markets. *Journal of Financial Economics*, 62, 247–292

Table 1
Liquidity, Correlation, and Error Correction Results

Panel A: ETF and S&P 500 Movements

	Correlation	Daily Error Correction Coefficient	Daily Error Correction *t*-statistic	1-Minute Error Correction Coefficient	1-Minute Error Correction *t*-statistic
		n = 2,409		*n* = 915,420	
SPY	0.98	-0.01	**-1.93**	-0.05	**-134.97**
IVV	0.99	-0.47	**-3.65**	-0.18	**-395.27**

Panel B: ETFs Movements

	Correlation	Daily Error Correction Coefficient	Daily Error Correction *t*-statistic	1-Minute Error Correction Coefficient	1-Minute Error Correction *t*-statistic
		n = 2,409		*n* = 915,420	
SPY	0.99	-0.05	**-2.17**	-0.01	**-39.52**
IVV		-0.04	**-1.88**	-0.20	**-419.83**

Panel C: ETF Liquidity and NAV deviations

	N	Min	25 P	Median	Mean	75 P	Max	Std. Dev
SPY Spread	2,409	0.000	0.009	0.009	0.017	0.014	0.429	0.020
SPY Value Traded	2,409	410	3,812	7,389	13,374	20,911	93,188	13,209
IVV Spread	2,409	0.009	0.016	0.023	0.046	0.056	0.664	0.048
IVV Value Traded	2,409	2	33	111	220	341	1,767	256
NAV Deviation	2,409	-0.364	-0.180	-0.113	-0.097	0.008	0.355	0.128

This table contains liquidity, correlation, and error correction results for the February 2001 – August 2010 period. Daily data in Panels A and B are from CRSP. Intraday data in these two panels are from the Thompson Reuters Tick History (TRTH) database. *t*-statistics that are statistically significant at the 10% level or higher are in bold. Panel C is based on daily CRSP data. Spreads, and NAV Deviation are in percent, while Value Traded is in US$ millions.

Table 2
Arbitrage Profits and Durations

Panel A: SPY Underpriced / IVV Overpriced - Profit and Duration Summary Statistics

	N	Min	25 Per	Median	Mean	75 Per	Max	Std Dev
Profit	90	0.20	0.23	0.27	0.33	0.34	1.87	0.22
Duration	90	0.32	0.76	2.27	86.46	21.25	3,769.24	404.95

Panel B: IVV Underpriced / SPY Overpriced - Profit and Duration Summary Statistics

	N	Min	25 Per	Median	Mean	75 Per	Max	Std Dev
Profit	93	0.20	0.24	0.27	0.32	0.33	1.27	0.15
Duration	93	0.06	0.52	0.92	3.92	1.64	186.96	19.57

Panel C: SPY Underpriced / IVV Overpriced - Mean Profit By Minute

Minute	-5	-4	-3	-2	-1	0	1	2
Mean	0.00	-0.02	-0.01	0.01	0.01	0.10	0.09	0.03
t-statistic	-0.34	-1.10	-0.51	0.45	0.95	**6.64**	**4.10**	1.25

Panel D: IVV Underpriced / SPY Overpriced - Mean Profit By Minute

Minute	-5	-4	-3	-2	-1	0	1	2
Mean	-0.06	-0.07	-0.05	-0.04	-0.01	0.08	0.06	0.00
t-statistic	**-5.71**	**-5.99**	**-4.69**	**-3.86**	-1.15	**5.65**	**3.79**	0.52

Panel E: Profit Per Annum

	N	Min	25 Per	Median	Mean	75 Per	Max	Std Dev
Profit		0.69	2.34	5.31	6.57	10.23	19.29	5.96

The data are from the Thompson Reuters Tick History (TRTH) database and the results relate to the February 2001 – August 2010 period. All profits are in percent and durations are in minutes. *t*-statistics that are statistically significant at the 10% level or higher are in bold.

Table 3
Determinants of Instances of Arbitrage and Arbitrage Profits

Panel A: Determinants of Instances of Arbitrage

	Buy SPY, Sell IVV, n = 90					*Buy IVV, Sell SPY, n = 93*				
Spread	0.14					0.25				
	5.00					**7.30**				
Value		0.07					-0.58			
		0.68					**-6.63**			
NAV Difference			-2.93					-9.05		
			-3.11					**-7.96**		
VIX				0.03					0.05	
				4.29					**8.78**	
S&P 500					-2.75					-23.94
					-0.22					**-2.43**
Constant	-3.98	-4.07	-3.09	-4.29	-3.45	-4.54	1.25	-2.59	-4.80	-3.48
	-23.70	**1.80**	**-19.99**	**-17.61**	**-29.19**	**-22.63**	**1.80**	**-19.28**	**-23.48**	**-27.68**
Adjusted R^2	0.04	0.00	0.01	0.02	0.00	0.13	0.04	0.08	0.06	0.01

Panel B: Determinants of Arbitrage Profits										
	Buy SPY, Sell IVV, n = 90					*Buy IVV, Sell SPY, n = 93*				
Spread	0.16					-0.21				
	0.64					-0.82				
Value		0.54					-5.43			
		0.36					**-1.71**			
NAV Difference			6.50					-23.66		
			0.42					-1.38		
VIX				0.02					-0.14	
				0.42					-1.10	
S&P 500					-0.56					0.36
					-0.70					0.69
Constant	26.22	22.47	26.30	26.29	26.98	29.95	70.75	29.58	29.58	28.47
	13.76	**1.84**	**13.37**	**13.37**	**20.40**	**9.85**	**2.72**	**13.39**	**13.39**	**17.51**
Adjusted R^2	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	0.06	0.00	0.00	-0.01

The spread and value data are from the Thompson Reuters Tick History (TRTH) database, NAV data are from the ETF provider websites and Morningstar, while VIX data are from Global Financial Data. The results relate to the February 2001 – August 2010 period. The Panel A results are based on a logit regression of a dependant variable that equals one on days an arbitrage opportunity is created and zero otherwise. Spread is average daily quoted spread across the ETFs (multiplied by 10,000), Value is the log of the average value traded, and NAV Difference is absolute value of the percentage difference in NAV. S&P 500 is the return on the S&P 500 index from CRSP. *z*-statistics (Panel A) and *t*-statistics (Panel B) are provided under the coefficients. Those that are statistically significant at the 10% level or more are in bold.

Table 4
Market Characteristics

Panel A: SPY Underpriced / IVV Overpriced						
	SPY			IVV		
	t-5 to *t*-2	*t*-1 to t+1	t+2 to t+5	*t*-5 to *t*-2	*t*-1 to t+1	t+2 to t+5
Spread	0.103	**0.413**	0.294	0.257	**0.508**	0.478
Depth	-0.074	-0.073	-0.042	-0.032	-0.006	-0.038
Order Imbalance	-0.073	**0.406**	**-0.061**	0.137	**0.167**	0.157
Return Std. Dev.	0.179	**1.317**	**0.704**	0.884	1.285	1.390
Trade Value	0.337	**2.305**	**1.805**	0.460	**0.929**	0.752
Spread Std. Dev.	0.125	**0.612**	0.625	1.686	**2.810**	**1.903**

Panel A: IVV Underpriced / SPY Overpriced						
	SPY			IVV		
	t-5 to *t*-2	*t*-1 to t+1	t+2 to t+5	*t*-5 to *t*-2	*t*-1 to t+1	t+2 to t+5
Spread	0.030	**0.176**	0.160	0.101	0.159	0.195
Depth	-0.046	-0.045	**-0.126**	-0.103	**-0.179**	**-0.067**
Order Imbalance	0.000	**0.271**	**-0.080**	0.154	0.172	0.169
Return Std. Dev.	0.067	**0.888**	**0.446**	1.786	1.827	1.375
Trade Value	0.076	**0.986**	1.120	0.708	**1.444**	**0.559**
Spread Std. Dev.	-0.012	**0.235**	0.184	1.090	**3.000**	2.080

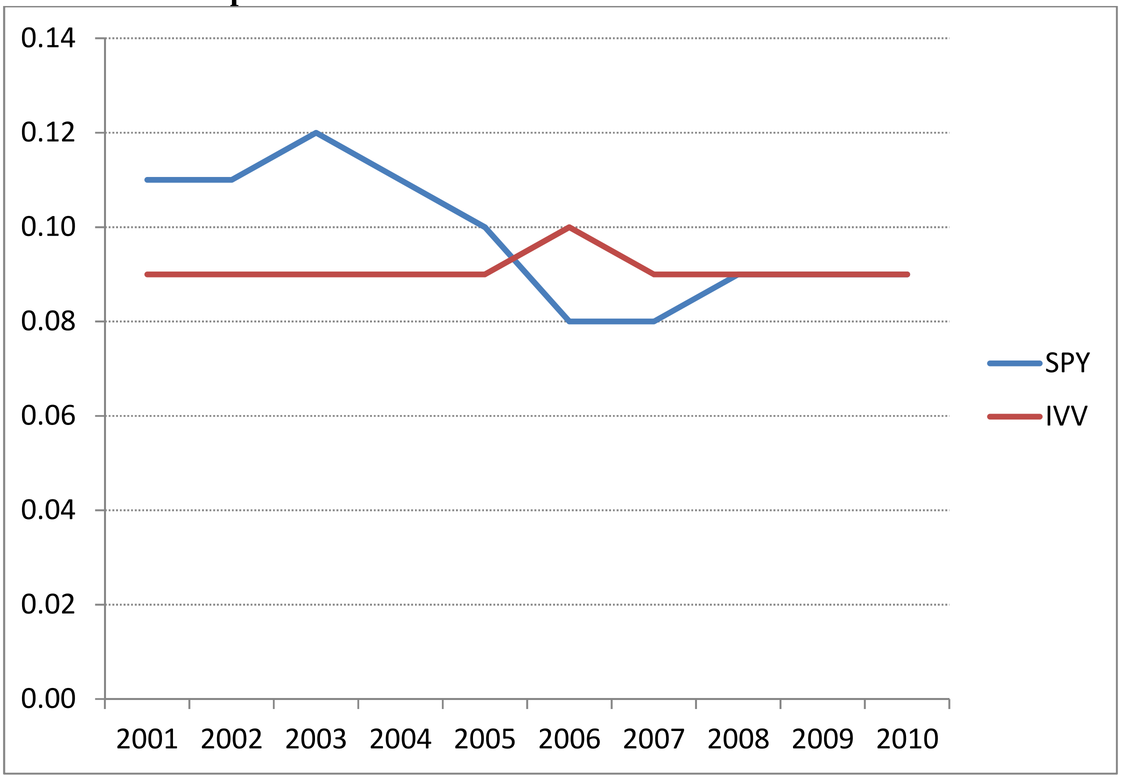
The data are from the Thompson Reuters Tick History (TRTH) database and the results relate to the February 2001 – August 2010 period. . Spread, depth, order imbalance, return standard deviation, trade value, and spread standard deviation are percentage increases or decreases based on the numbers from the same time of the day on the previous 20 trading days. The statistical significance of the difference of the variables across adjacent windows is calculated using a bootstrap test. Statistical significance at the 10% level or higher is in bold on the column to the right. Therefore bold in column *t*-1 to t+1 means these numbers are statistically significantly different to those in *t*-5 to *t*-2. Bold in column t+2 to t+5 means these numbers are statistically significantly different to those in *t*-1 to t+1.

Table 5
Convergence

		SPY Underpriced / IVV Overpriced $n = 90$			IVV Underpriced / SPY Overpriced $n = 93$		
		Number	Volume	$ Volume	Number	Volume	$ Volume
SPY	Buy	55%	54%	54%	44%	42%	42%
	Sell	45%	46%	46%	56%	58%	58%
	t-statistic	1.24	1.06	1.07	**-1.64**	**-2.31**	**-2.30**
IVV	Buy	50%	48%	48%	63%	58%	58%
	Sell	50%	52%	52%	37%	42%	42%
	t-statistic	-0.03	-0.63	-0.62	**3.54**	**2.28**	**2.29**
SPY	Aggressive Buy	50%	51%	51%	34%	36%	36%
	Aggressive Sell	41%	43%	43%	46%	52%	52%
	t-statistic	1.25	1.09	1.10	**-1.71**	**-2.18**	**-2.17**
IVV	Aggressive Buy	31%	32%	32%	36%	38%	38%
	Aggressive Sell	40%	41%	41%	22%	25%	25%
	t-statistic	-1.30	-1.36	-1.36	**2.08**	**1.89**	**1.89**

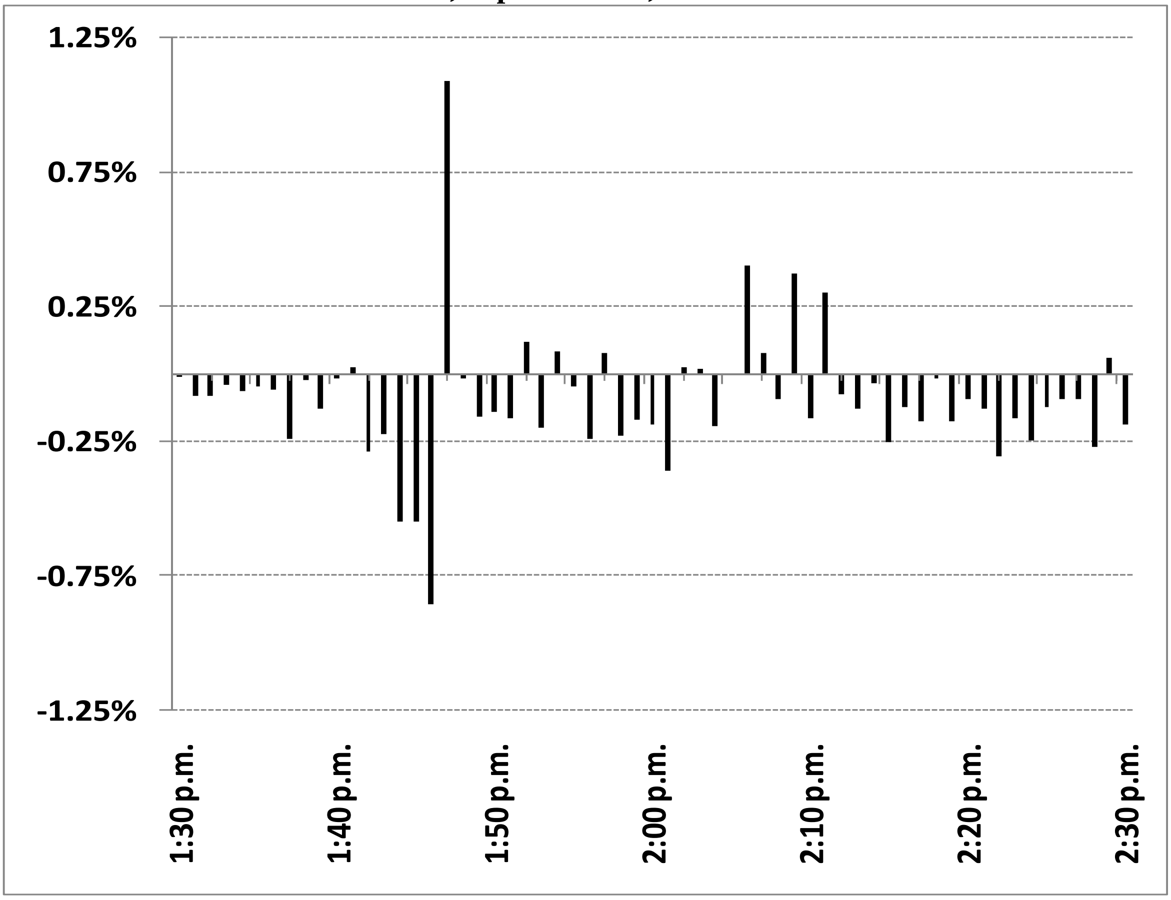
The data are from the Thompson Reuters Tick History (TRTH) database and the results relate to the February 2001 – August 2010 period. . Trades are classified based on the Lee and Ready (1991) algorithm. The proportion of buyer- and seller-initiated trades in the time between mispricing allowing arbitrage profits and their removal are presented. Aggressive buy (sell) trades are those trades at or above the prevailing ask price (at or below the prevailing bid price). Statistically significant differences between buy and sell and between aggressive buy and aggressive sell trades (at the 10% level are in bold).

Figure 1
SPY and IVV Expense Ratios



Net expense ratios for the SPY and IVV from Morningstar.

Figure 2
IVV Deviations from the S&P 500, September 29, 2008



Deviations from parity between the IVV and S&P 500 during one-minute intervals on 29 September 2008. Numbers greater (less) than zero indicate the IVV is trading at a premium (discount).

Figure 3a
Monthly SPY and IVV Trading Value versus Median Stock

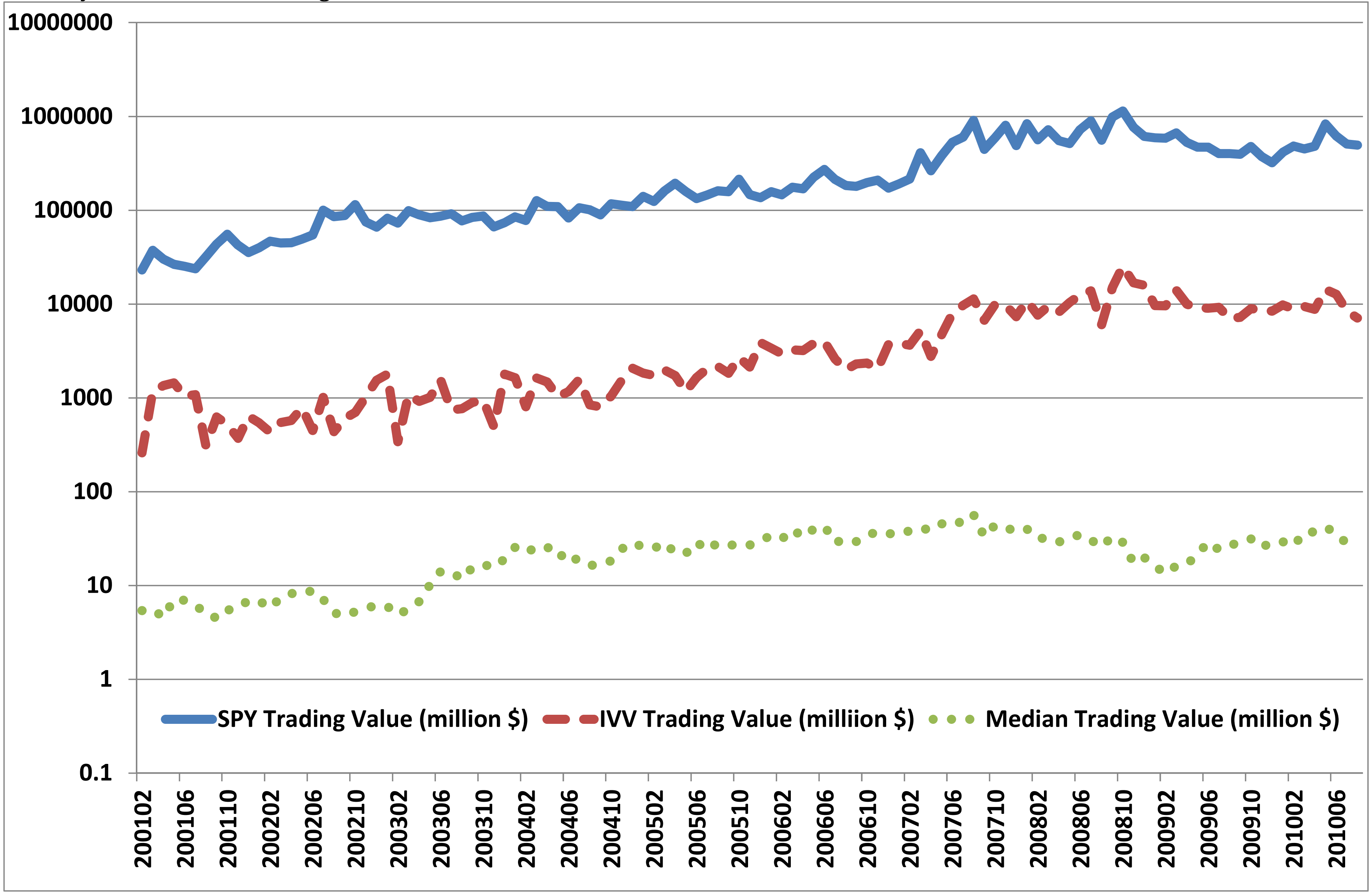


Figure 3b
SPY and IVV Spread versus Median Stock

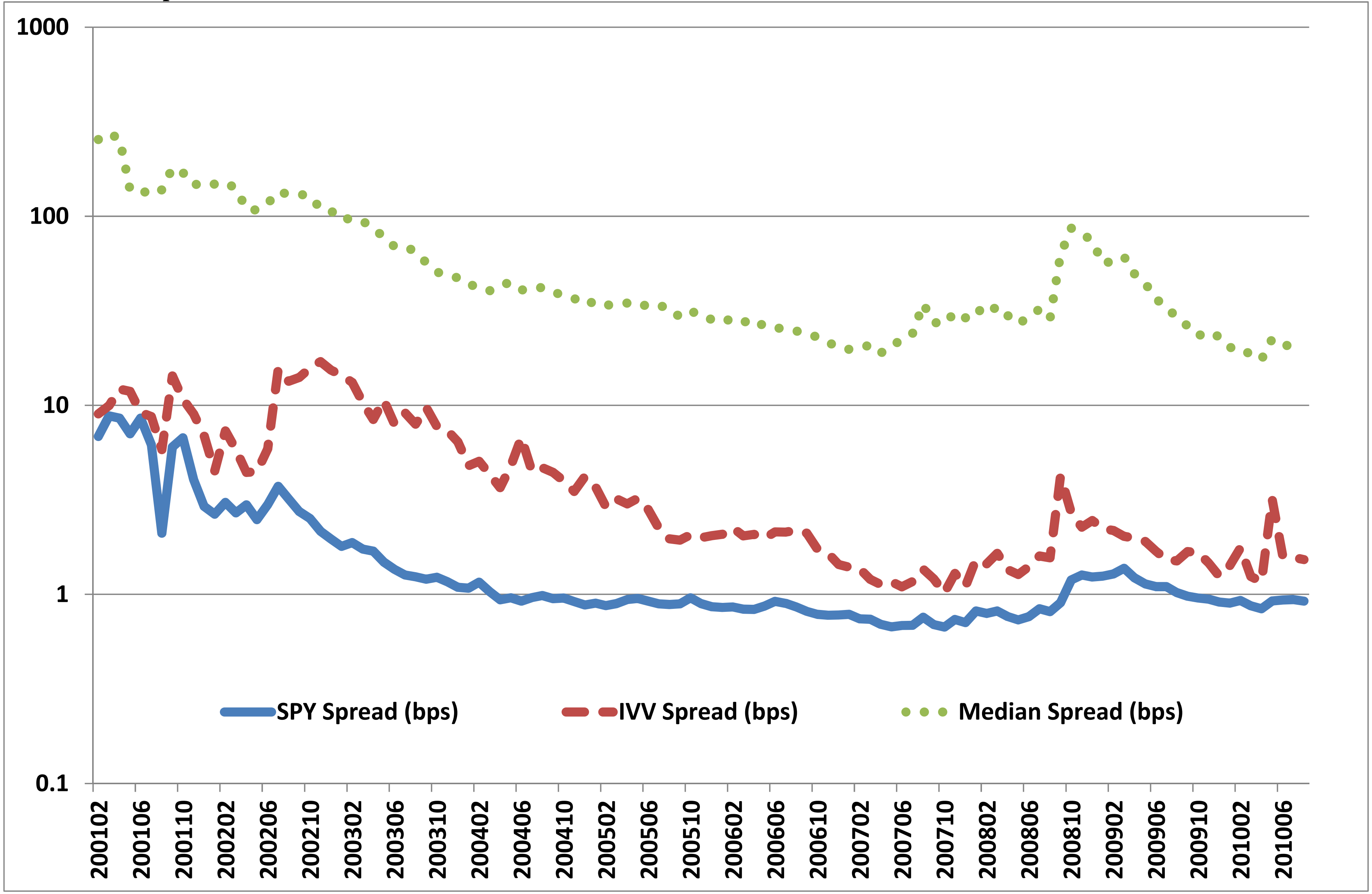
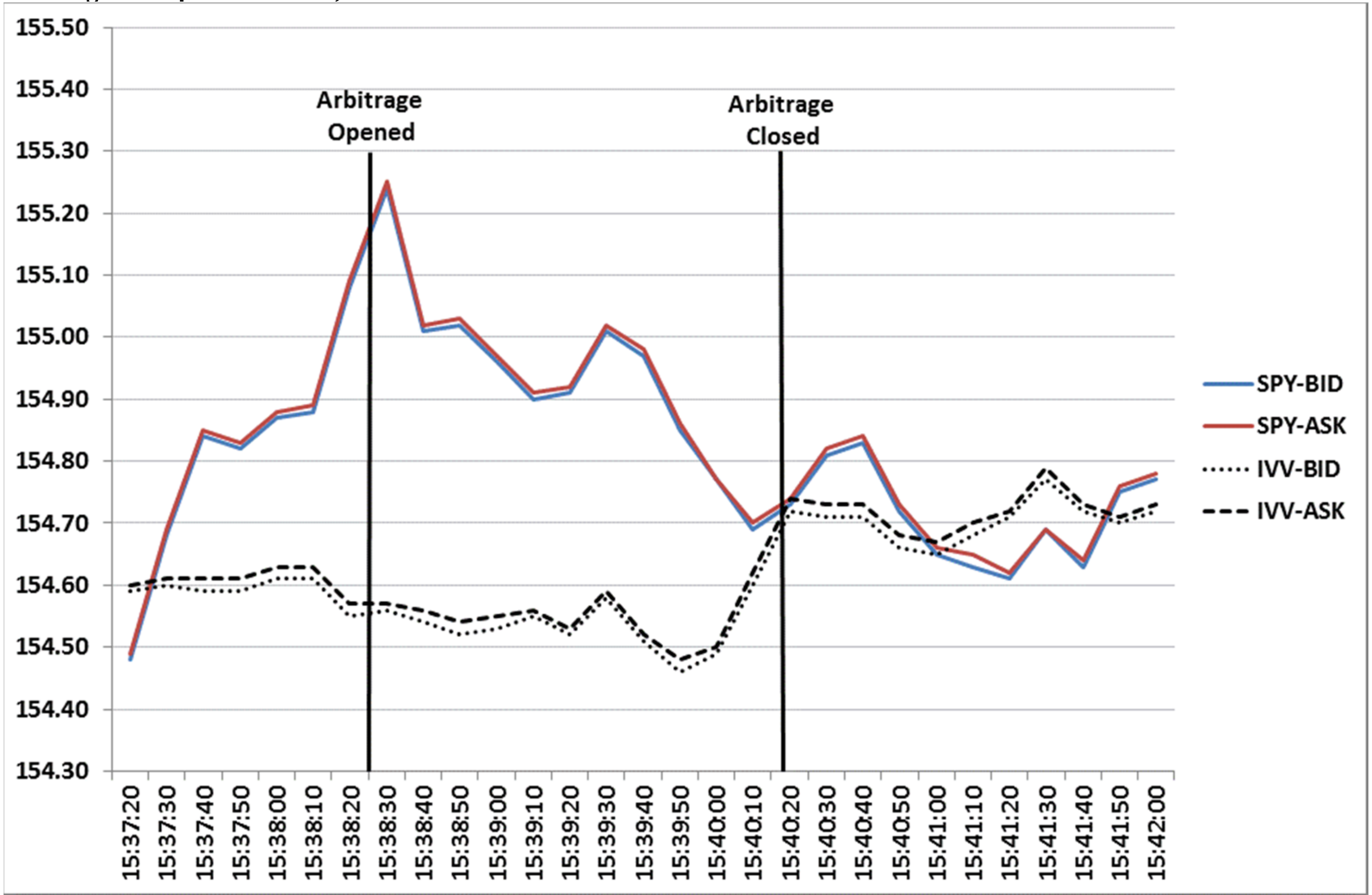
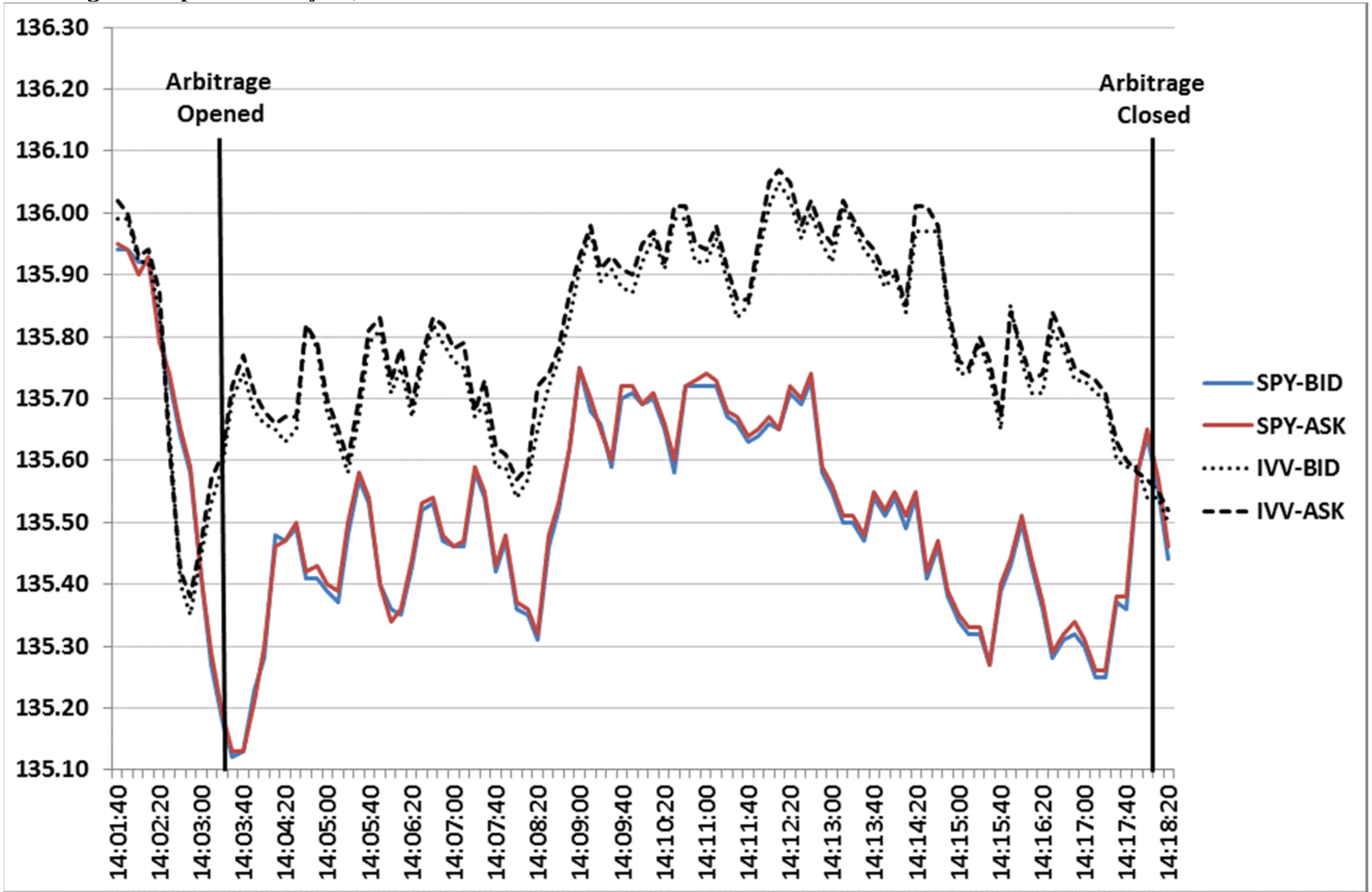


Figure 4a
Arbitrage Example October 31, 2007



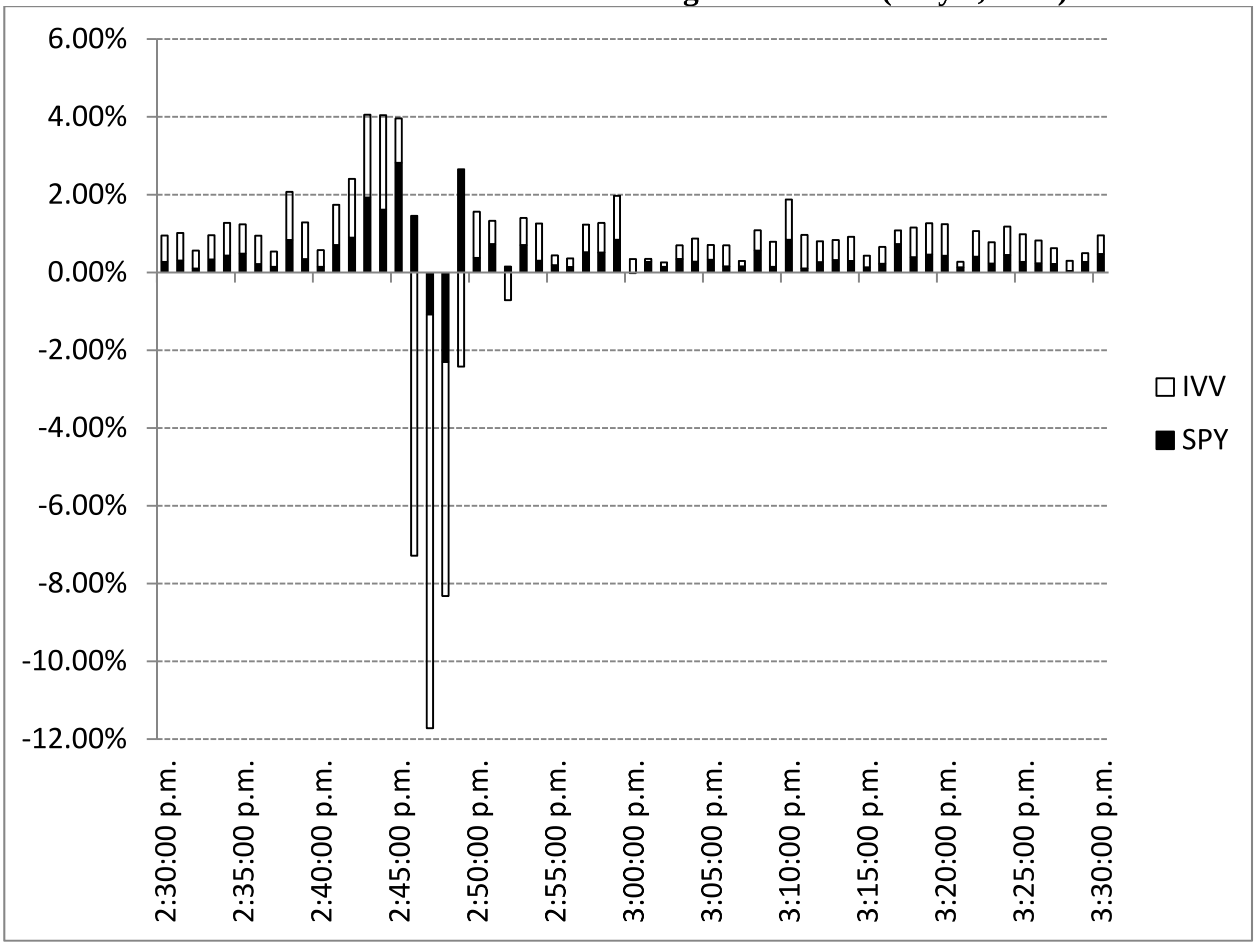
An arbitrage that was opened by selling (buying) the SPY (IVV) at the bid (ask) and closed at by selling (buying) the IVV (SPY) at the bid (ask).

Figure 4b
Arbitrage Example February 20, 2008



An arbitrage that was opened by selling (buying) the IVV (SPY) at the bid (ask) and closed at by selling (buying) the SPY (IVV) at the bid (ask).

Figure 5
IVV and SPY Deviations from the S&P 500 During Flash Crash (May 6, 2010)



Deviations from parity between the IVV and S&P 500 and SPY and S&P 500 during one-minute intervals during the Flash Crash on May 6, 2010. Numbers greater (less) than zero indicate the IVV is trading at a premium (discount).

Appendix 1
Arbitrage Profits for Alternative Strategies

Threshold %	N	Min	Median	Mean	Max	Std Dev
Panel A: SPY Underpriced / IVV Overpriced						
0.10 / 0.00	371	0.10	0.15	0.18	1.66	0.12
0.15 / 0.00	164	0.15	0.21	0.25	1.66	0.16
0.30 / -0.10	63	0.21	0.32	0.36	2.06	0.26
0.40 / -0.20	38	0.21	0.32	0.37	0.91	0.15
Panel B: IVV Underpriced / SPY Overpriced						
0.10 / 0.00	517	0.10	0.15	0.18	1.32	0.11
0.15 / 0.00	210	0.15	0.22	0.26	1.32	0.15
0.30 / -0.10	29	0.22	0.34	0.42	1.19	0.21
0.40 / -0.20	14	0.33	0.45	0.54	1.19	0.24

Profit summary statistics for strategies that require 0.3% (0.4%) profit for the arbitrage to be opened and are closed at a loss of 0.1% (0.2) to ensure an overall profit of 0.2%.

Appendix 2
Median Arbitrage Profits by Year

	SPY Underpriced / IVV Overpriced		IVV Underpriced / SPY Overpriced	
	N	Profit	N	Profit
2001	11	0.27	18	0.27
2002	17	0.33	44	0.28
2003	16	0.25	20	0.26
2004	7	0.22	3	0.24
2005	1	0.24	2	0.23
2006	8	0.26	-	-
2007	13	0.25	4	0.24
2008	16	0.30	1	0.51
2009	1	1.87	1	0.50

The data are from the Thompson Reuters Tick History (TRTH) database and the results relate to the February 2001 – August 2010 period. All profits are in percent.

Appendix 3
Control Variable Summary Statistics

	N	Min	25 Per	Median	Mean	75 Per	Max	Std Dev
VIX	2,409	9.89	14.59	21.22	22.85	27.18	80.86	10.70
S&P 500	2,409	-9.03	-0.61	0.06	0.00	0.62	11.58	1.39

Summary statistics for the control variables not presented in Table 1 Panel C. Numbers other than N are expressed as a percentage.