ETF Arbitrage and Return Predictability

David C. Brown
University of Arizona

Shaun William Davies
University of Colorado Boulder

Matthew Ringgenberg
University of Utah

January 5, 2018

American Finance Association Annual Meeting
Motivation

Demand Shocks and Absolute Price Efficiency

- Demand shocks hit assets and move prices
  - Informed traders (Kyle 1985)
  - Noise traders (Shleifer and Summers 1990)

Sources of demand shocks are often unknown for long periods of time, leading to predictable returns. Fire sales (Coval and Stafford 2007) and mutual fund flows (Lou 2012) are examples of such events.
Motivation

Demand Shocks and Absolute Price Efficiency

- Demand shocks hit assets and move prices
  - Informed traders (Kyle 1985)
  - Noise traders (Shleifer and Summers 1990)

- Sources of demand shocks are often unknown for long periods of time, leading to predictable returns
  - Fire sales (Coval and Stafford 2007)
  - Mutual fund flows (Lou 2012)
**Motivation**

**Demand Shocks and Absolute Price Efficiency**

- Demand shocks hit assets and move prices
  - Informed traders (Kyle 1985)
  - Noise traders (Shleifer and Summers 1990)
- Sources of demand shocks are often unknown for long periods of time, leading to predictable returns
  - Fire sales (Coval and Stafford 2007)
  - Mutual fund flows (Lou 2012)
- Thus, demand shocks often result in **absolute** price inefficiency
Relative Price Efficiency and ETFs

- When identical assets exist, arbitrageurs ensure the law of one price holds
Motivation

Relative Price Efficiency and ETFs

- When identical assets exist, arbitrageurs ensure the law of one price holds
  - For example, ETFs and their underlying securities (NAV)
Relative Price Efficiency and ETFs

- When identical assets exist, arbitrageurs ensure the law of one price holds
  - For example, ETFs and their underlying securities (NAV)
  - Authorized participants make arbitrage trades to maintain relative price efficiency (Petajisto 2017, Engle and Sarkar 2006)
Relative Price Efficiency and ETFs

- When identical assets exist, arbitrageurs ensure the law of one price holds
  - For example, ETFs and their underlying securities (NAV)
- Authorized participants make arbitrage trades to maintain relative price efficiency (Petajisto 2017, Engle and Sarkar 2006)
- Relative price efficiency does not imply absolute price efficiency
ETF Arbitrage Example

Non-Fundamental Demand Shocks and Arbitrage Trades
ETF Share Price and Underlying NAV

ETF Premium

Brown, Davies and Ringgenberg
ETF Arbitrage Example

Non-Fundamental Demand Shocks and Arbitrage Trades
ETF Share Price and Underlying NAV

ETF_0
NAV_0
ETF_1
NAV_1
Relative Demand Shocks
ETF Premium

Brown, Davies and Ringgenberg
ETF Arbitrage and Return Predictability
ETF Arbitrage Example

Non-Fundamental Demand Shocks and Arbitrage Trades
ETF Share Price and Underlying NAV

ETF Arbitrage and Return Predictability

Brown, Davies and Ringgenberg
ETF Arbitrage Example

Non-Fundamental Demand Shocks and Arbitrage Trades
ETF Share Price and Underlying NAV

ETF Arbitrage and Return Predictability

Brown, Davies and Ringgenberg
ETF Arbitrage Example

Fundamental Demand Shocks and Arbitrage Trades
ETF Share Price and Underlying NAV

ETF\textsubscript{0}
NAV\textsubscript{0}

Brown, Davies and Ringgenberg
Motivation

ETF Arbitrage Example

Fundamental Demand Shocks and Arbitrage Trades

ETF Share Price and Underlying NAV

$ETF_0$, $NAV_0$, $ETF_1$, $NAV_1$
ETF Arbitrage Example

Fundamental Demand Shocks and Arbitrage Trades
ETF Share Price and Underlying NAV

ETF $0$
NAV $0$

ETF $1$
NAV $1$

ETF $2$
NAV $2$

Relative Demand Shocks
Arbitrage Activity

Brown, Davies and Ringgenberg
ETF Arbitrage Example

Motivation

ETF Arbitrage and Return Predictability

Brown, Davies and Ringgenberg
Null Hypothesis: Weak-Form Market Efficiency

- Relative demand shocks lead to arbitrage activity
Motivation

Null Hypothesis: Weak-Form Market Efficiency

- Relative demand shocks lead to arbitrage activity
- Following arbitrage activity, prices should return to fundamental values
  - Non-fundamental shocks $\rightarrow$ price reversions
  - Fundamental shocks $\rightarrow$ price continuation
Motivation

Null Hypothesis: Weak-Form Market Efficiency

- Relative demand shocks lead to arbitrage activity
- Following arbitrage activity, prices should return to fundamental values
  - Non-fundamental shocks $\rightarrow$ price reversions
  - Fundamental shocks $\rightarrow$ price continuation
- Arbitrage activity is:
  - symptomatic of relative demand shocks
Motivation

Null Hypothesis: Weak-Form Market Efficiency

- Relative demand shocks lead to arbitrage activity
- Following arbitrage activity, prices should return to fundamental values
  - Non-fundamental shocks $\rightarrow$ price reversion
  - Fundamental shocks $\rightarrow$ price continuation
- Arbitrage activity is:
  1. symptomatic of relative demand shocks
  2. observable
Motivation

Null Hypothesis: Weak-Form Market Efficiency

- Relative demand shocks lead to arbitrage activity
- Following arbitrage activity, prices should return to fundamental values
  - Non-fundamental shocks $\rightarrow$ price reversions
  - Fundamental shocks $\rightarrow$ price continuation
- Arbitrage activity is:
  1. symptomatic of relative demand shocks
  2. observable
- Absolute price efficiency should be quickly restored
Motivation

Null Hypothesis: Weak-Form Market Efficiency

- Relative demand shocks lead to arbitrage activity
- Following arbitrage activity, prices should return to fundamental values
  - Non-fundamental shocks $\rightarrow$ price reversions
  - Fundamental shocks $\rightarrow$ price continuation
- Arbitrage activity is:
  1. symptomatic of relative demand shocks
  2. observable
- Absolute price efficiency should be quickly restored
- Null hypothesis: Monthly arbitrage activity does not predict monthly returns
Motivation

What We Do

Overview

- Use ETF creation / redemption mechanism to test whether markets incorporate the information in arbitrage trades.

Preview of Results

- Arbitrage activity predicts future asset returns for both the underlying stocks and ETFs themselves.
- Arbitrage activity is associated with return reversals.
- ETF investors collectively mistime the market.
Motivation

What We Do

Overview

- Use ETF creation/redemption mechanism to test whether markets incorporate the information in arbitrage trades
- ETFs provide a unique opportunity to identify demand shocks
  - Authorized Participants engage in arbitrage trades to correct mispricing from relative demand shocks
  - Daily share changes provide an observable measure of arbitrage activity

Preview of Results

- Arbitrage activity predicts future asset returns
  - For both the underlying stocks and ETFs themselves
  - Arbitrage activity is associated with return reversals
  - ETF investors collectively mistime the market
Motivation

What We Do

Overview

- Use ETF creation / redemption mechanism to test whether markets incorporate the information in arbitrage trades
- ETFs provide a unique opportunity to identify demand shocks
  - Authorized Participants engage in arbitrage trades to correct mispricing from relative demand shocks
  - Daily share changes provide an observable measure of arbitrage activity

Preview of Results

- Arbitrage activity predicts future asset returns
  - For both the underlying stocks and ETFs themselves
Motivation

What We Do

Overview

- Use ETF creation / redemption mechanism to test whether markets incorporate the information in arbitrage trades
- ETFs provide a unique opportunity to identify demand shocks
  - Authorized Participants engage in arbitrage trades to correct mispricing from relative demand shocks
  - Daily share changes provide an observable measure of arbitrage activity

Preview of Results

- Arbitrage activity predicts future asset returns
  - For both the underlying stocks and ETFs themselves
- Arbitrage activity is associated with return reversals
Motivation

What We Do

Overview

- Use ETF creation / redemption mechanism to test whether markets incorporate the information in arbitrage trades
- ETFs provide a unique opportunity to identify demand shocks
  - Authorized Participants engage in arbitrage trades to correct mispricing from relative demand shocks
  - Daily share changes provide an observable measure of arbitrage activity

Preview of Results

- Arbitrage activity predicts future asset returns
  - For both the underlying stocks and ETFs themselves
- Arbitrage activity is associated with return reversals
- ETF investors collectively mistime the market
Monthly data for 2,196 ETFs spanning 2007 to 2016
ETF Sample

- Monthly data for 2,196 ETFs spanning 2007 to 2016
Empirical Analysis: Data

ETF Sample

- Monthly data for 2,196 ETFs spanning 2007 to 2016

ETFs “mature” once creation/redemption activity exceeds 50% of days
Empirical Analysis: ETF-Level Evidence

Return Predictability Methodology

- Sort ETFs into deciles based on net creations/redemptions over past month
Empirical Analysis: ETF-Level Evidence

Return Predictability Methodology

- Sort ETFs into deciles based on net creations/redemptions over past month
- Analyze differences in portfolio returns between high redemption (Decile 1) and high creation (Decile 10) ETFs

Regress monthly ETF returns on factors (raw returns, 3-factor, 4-factor and 5-factor models)
Consistent results using NAV returns
Consistent results for stock-level returns using aggregated ETF creations and redemptions

Brown, Davies and Ringgenberg
Empirical Analysis: ETF-Level Evidence

Return Predictability Methodology

- Sort ETFs into deciles based on net creations/redemptions over past month
- Analyze differences in portfolio returns between high redemption (Decile 1) and high creation (Decile 10) ETFs
- Regress monthly ETF returns on factors (raw returns, 3-factor, 4-factor and 5-factor models)

Consistent results using NAV returns
Consistent results for stock-level returns using aggregated ETF creations and redemptions
Empirical Analysis: ETF-Level Evidence

Return Predictability Methodology

- Sort ETFs into deciles based on net creations/redemptions over past month
- Analyze differences in portfolio returns between high redemption (Decile 1) and high creation (Decile 10) ETFs
- Regress monthly ETF returns on factors (raw returns, 3-factor, 4-factor and 5-factor models)
  - Consistent results using NAV returns
Empirical Analysis: ETF-Level Evidence

Return Predictability Methodology

- Sort ETFs into deciles based on net creations/redemptions over past month
- Analyze differences in portfolio returns between high redemption (Decile 1) and high creation (Decile 10) ETFs
- Regress monthly ETF returns on factors (raw returns, 3-factor, 4-factor and 5-factor models)
  - Consistent results using NAV returns
  - Consistent results for stock-level returns using aggregated ETF creations and redemptions
Empirical Analysis: ETF-Level Evidence

ETF Arbitrage Negatively Predicts Returns

High Redemption vs. High Creation Raw ETF Returns

<table>
<thead>
<tr>
<th>Monthly Return (%)</th>
<th>Equal-Weighted (1.99%*<strong>), Value-Weighted (1.20%</strong>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redemptions (Decile 1)</td>
<td>0.681** (Redemptions), 0.712* (Creation), -1.312*** (Redemptions), -0.485 (Creation)</td>
</tr>
<tr>
<td>Creations (Decile 10)</td>
<td>0.712* (Redemptions), -0.485 (Creation)</td>
</tr>
</tbody>
</table>

ETF Arbitrage and Return Predictability

Brown, Davies and Ringgenberg
Empirical Analysis: ETF-Level Evidence

ETF Arbitrage Negatively Predicts Returns

High Redemption vs. High Creation Raw ETF Returns

Equal-weighted → 26.7% annualized raw return
Empirical Analysis: ETF-Level Evidence

ETF Arbitrage Negatively Predicts Returns

High Redemption vs. High Creation Raw ETF Returns

<table>
<thead>
<tr>
<th></th>
<th>Monthly Return (%)</th>
<th>Equal-Weighted (1.99%***)</th>
<th>Value-Weighted (1.20%**)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redemption</td>
<td>0.681**</td>
<td>-1.312***</td>
<td>0.712*</td>
</tr>
<tr>
<td>Creation</td>
<td>-0.485</td>
<td></td>
<td>-0.485</td>
</tr>
</tbody>
</table>

Value-weighted $\rightarrow$ 15.4% annualized raw return
Return reversion suggests relative demand shocks are non-fundamental, consistent with Ben-David, Franzoni, Moussawi (Forthcoming JF)
Empirical Analysis: ETF-Level Evidence

ETF Arbitrage Negatively Predicts Returns

High Redemption vs. High Creation Raw ETF Returns

<table>
<thead>
<tr>
<th></th>
<th>Equal-Weighted (1.99%***)</th>
<th>Value-Weighted (1.20%**)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redemptions (Decile 1)</td>
<td>0.681**</td>
<td>0.712*</td>
</tr>
<tr>
<td>Creations (Decile 10)</td>
<td>-1.312***</td>
<td>-0.485</td>
</tr>
</tbody>
</table>

Similar results using factor-based alphas or NAVs
Empirical Analysis: ETF-Level Evidence

Predictability Stronger in High-Activity ETFs

High Redemption vs. High Creation Raw ETF Returns by ETF Activity Terciles

<table>
<thead>
<tr>
<th>Activity Tercile</th>
<th>Redemptions (Decile 1)</th>
<th>Creations (Decile 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Activity (0.10%)</td>
<td>0.62</td>
<td>0.52</td>
</tr>
<tr>
<td>Medium Activity (1.50%**)</td>
<td>0.86**</td>
<td>-0.64</td>
</tr>
<tr>
<td>High Activity (1.83%**)</td>
<td>1.04**</td>
<td>-0.79</td>
</tr>
</tbody>
</table>

Brown, Davies and Ringgenberg
Empirical Analysis: ETF-Level Evidence

Predictability Stronger in High-Activity ETFs

High Redemption vs. High Creation Raw ETF Returns by ETF Activity Terciles

More arbitrage activity is associated with more return predictability

Brown, Davies and Ringgenberg

ETF Arbitrage and Return Predictability
Empirical Analysis: ETF-Level Evidence

Results Concentrated in Levered and Broad-Market ETFs

High Redemption vs. High Creation Raw ETF Returns by ETF Category

Montly Return (%)

-4.00  -3.00  -2.00  -1.00  0.00  1.00  2.00

-4.00  -3.00  -2.00  -1.00  0.00  1.00  2.00

Overall (1.56%***)
Levered (4.23%***)
Broad Market (3.67%***)
Sector-Based (0.37%)
Bond (-0.22%)
Commodity (0.93%)
International (0.35%)

Redemptions (Decile 1) | Creations (Decile 10)

Brown, Davies and Ringgenberg

ETF Arbitrage and Return Predictability
Empirical Analysis: ETF-Level Evidence

Results Concentrated in Levered and Broad-Market ETFs

High Redemption vs. High Creation Raw ETF Returns by ETF Category

Levered ETFs show the strongest predictability
Empirical Analysis: ETF-Level Evidence

Results Concentrated in Levered and Broad-Market ETFs

High Redemption vs. High Creation Raw ETF Returns by ETF Category

Broad market ETFs, not niche ETFs, drive our results
Empirical Analysis: Time Series Evidence

What Does This Cost Investors?

- Our results suggest ETF investors collectively mistime market
  - ETF creations $\rightarrow$ lower future ETF performance
  - ETF redemptions $\rightarrow$ higher future ETF performance

Implication: investors consistently overpay to gain ETF exposure

Individual cost depends on frequency of trade

We consider a representative investor who re-balances according to creations/redemptions

Brown, Davies and Ringgenberg

ETF Arbitrage and Return Predictability
Empirical Analysis: Time Series Evidence

What Does This Cost Investors?

- Our results suggest ETF investors collectively mistime market
  - ETF creations $\rightarrow$ lower future ETF performance
  - ETF redemptions $\rightarrow$ higher future ETF performance
  - Implication: investors consistently overpay to gain ETF exposure
Our results suggest ETF investors collectively mistime market

- ETF creations $\rightarrow$ lower future ETF performance
- ETF redemptions $\rightarrow$ higher future ETF performance

Implication: investors consistently overpay to gain ETF exposure

Individual cost depends on frequency of trade
Empirical Analysis: Time Series Evidence

What Does This Cost Investors?

- Our results suggest ETF investors collectively mistime market
  - ETF creations $\rightarrow$ lower future ETF performance
  - ETF redemptions $\rightarrow$ higher future ETF performance
    - Implication: investors consistently overpay to gain ETF exposure
- Individual cost depends on frequency of trade
- We consider a representative investor who re-balances according to creations/redemptions
Time-Series Methodology

- Standard time-series analysis assumes fixed quantities of shares
Time-Series Methodology

- Standard time-series analysis assumes fixed quantities of shares
- ETF time-series analysis must account for creations and redemptions

Efficient fees capture the difference between actual and asset-weighted returns. We randomize ETF flows using block-bootstrap Monte Carlo methods to generate test statistics (p-values based on 1,000,000 simulations) and control for growth of the ETF industry over time.
Empirical Analysis: Time Series Evidence

Time-Series Methodology

- Standard time-series analysis assumes fixed quantities of shares
- ETF time-series analysis must account for creations and redemptions
- We generate share-growth-adjusted (i.e. asset-weighted) returns to account for total capital invested in ETFs
Time-Series Methodology

- Standard time-series analysis assumes fixed quantities of shares
- ETF time-series analysis must account for creations and redemptions
- We generate share-growth-adjusted (i.e. asset-weighted) returns to account for total capital invested in ETFs
- Effective fees capture difference between actual and asset-weighted returns
Empirical Analysis: Time Series Evidence

Time-Series Methodology

- Standard time-series analysis assumes fixed quantities of shares
- ETF time-series analysis must account for creations and redemptions
- We generate share-growth-adjusted (i.e. asset-weighted) returns to account for total capital invested in ETFs
- Effective fees capture difference between actual and asset-weighted returns
- We randomize ETF flows using block-bootstrap Monte Carlo methods to:
  - Generate test statistics (p-values based on 1,000,000 simulations)
  - Control for growth of ETF industry over time
Empirical Analysis: Time Series Evidence

Effective Fees Are More Negative Than Positive

Distribution of Effective Fee P-Values

Brown, Davies and Ringgenberg
ETF Arbitrage and Return Predictability
Empirical Analysis: Time Series Evidence

Effective Fees Are More Negative Than Positive

Distribution of Effective Fee P-Values

Equal-weighted → 12% < 0.05 p-value threshold
Empirical Analysis: Time Series Evidence

Effective Fees Are More Negative Than Positive

Distribution of Effective Fee P-Values

Value-weighted → 26% < 0.05 p-value threshold
Empirical Analysis: Time Series Evidence

Negative Effective Fee Examples: SPY & Total ETF AUM

- SPY (largest ETF, replicates S&P500):
  - Actual annual return (2007–2016): 6.89%
Empirical Analysis: Time Series Evidence

Negative Effective Fee Examples: SPY & Total ETF AUM

- SPY (largest ETF, replicates S&P500):
  - Actual annual return (2007–2016): 6.89%
  - Average simulated share-growth-adjusted annual return: 6.92%
Negative Effective Fee Examples: SPY & Total ETF AUM

- **SPY (largest ETF, replicates S&P500):**
  - Actual annual return (2007–2016): 6.89%
  - Average simulated share-growth-adjusted annual return: 6.92%
  - Realized share-growth-adjusted annual return: 5.44%
Empirical Analysis: Time Series Evidence

Negative Effective Fee Examples: SPY & Total ETF AUM

- **SPY (largest ETF, replicates S&P500):**
  - Actual annual return (2007–2016): 6.89%
  - Average simulated share-growth-adjusted annual return: 6.92%
  - Realized share-growth-adjusted annual return: 5.44%
  - Annualized Effective Fee: **1.48%**

Total ETF AUM (Aggregated)
- Annualized effective fee (2007–2016): 0.33%
- Annualized effective fee (2007–2011): 0.55%
- Annualized effective fee (2012–2016): 0.07%
- 0.07% on $2.3 trillion AUM → $1.6 billion of underperformance in 2016
Empirical Analysis: Time Series Evidence

Negative Effective Fee Examples: SPY & Total ETF AUM

- **SPY (largest ETF, replicates S&P500):**
  - Actual annual return (2007–2016): 6.89%
  - Average simulated share-growth-adjusted annual return: 6.92%
  - Realized share-growth-adjusted annual return: 5.44%
  - Annualized Effective Fee: **1.48%**

- **Total ETF AUM (Aggregated)**
  - Annualized effective fee (2007–2016): 0.33%
Empirical Analysis: Time Series Evidence

Negative Effective Fee Examples: SPY & Total ETF AUM

- **SPY (largest ETF, replicates S&P500):**
  - Actual annual return (2007–2016): 6.89%
  - Average simulated share-growth-adjusted annual return: 6.92%
  - Realized share-growth-adjusted annual return: 5.44%
  - Annualized Effective Fee: **1.48%**

- **Total ETF AUM (Aggregated)**
  - Annualized effective fee (2007–2016): 0.33%
  - Annualized effective fee (2007–2011): 0.55%
Empirical Analysis: Time Series Evidence

Negative Effective Fee Examples: SPY & Total ETF AUM

- **SPY (largest ETF, replicates S&P500):**
  - Actual annual return (2007–2016): 6.89%
  - Average simulated share-growth-adjusted annual return: 6.92%
  - Realized share-growth-adjusted annual return: 5.44%
  - Annualized Effective Fee: **1.48%**

- **Total ETF AUM (Aggregated)**
  - Annualized effective fee (2007–2016): 0.33%
  - Annualized effective fee (2007–2011): 0.55%
  - Annualized effective fee (2012–2016): 0.07%
Empirical Analysis: Time Series Evidence

Negative Effective Fee Examples: SPY & Total ETF AUM

- **SPY (largest ETF, replicates S&P500):**
  - Actual annual return (2007–2016): 6.89%
  - Average simulated share-growth-adjusted annual return: 6.92%
  - Realized share-growth-adjusted annual return: 5.44%
  - Annualized Effective Fee: **1.48%**

- **Total ETF AUM (Aggregated)**
  - Annualized effective fee (2007–2016): 0.33%
  - Annualized effective fee (2007–2011): 0.55%
  - Annualized effective fee (2012–2016): 0.07%
  - 0.07% on $2.3 trillion AUM $\rightarrow$ **$1.6 billion** of underperformance in 2016
ETF arbitrage activity negatively predicts future returns
Conclusion

Take Aways

1. ETF arbitrage activity negatively predicts future returns

2. Observable, non-fundamental demand shocks are not quickly offset by market participants
Take Aways

1. ETF arbitrage activity negatively predicts future returns

2. Observable, non-fundamental demand shocks are not quickly offset by market participants

3. Information conveyed by arbitrage trades is not fully incorporated into prices