Caught On Tape: Predicting Institutional Ownership With Order Flow

John Campbell, Tarun Ramadorai and Tuomo Vuolteenaho

Harvard University and University of Oxford

Broad Research Agenda: Equity Flows of Institutional Investors

- Do institutional investors respond differently to information than individuals?
- Do institutions take liquidity or provide it? Do they stabilize or destabilize prices?
- Need information on returns, high frequency institutional equity flows, and high frequency cash- flow relevant information.
- But high frequency institutional equity flows are not readily available! (Few sources such as Plexus, TORQ, NYSE Audit Trail: limited coverage)
- Need to construct…

Existing Studies on Institutional Trading

- At quarterly frequency, institutional holdings are readily measured from 13-F filings (Spectrum data)
- Intriguing results:
  - Stock returns and institutional flows are contemporaneously correlated (Grinblatt, Titman, and Wermers 1995, Wermers 1999, 2000, Nofsinger and Sias 1999)
  - Institutional investors buy shares from individuals in response to good cash-flow news (Cohen, Gompers and Vuolteenaho 2002)
- We would like to know if these results hold up at higher frequency

Existing Studies on Institutional Trading

- High-frequency proprietary data
  - State Street custodial data (Froot, O'Connell, Seasholes 2001, Froot and Ramadorai 2005a, Froot and Teo 2005)
  - NYSE Audit Trail (Kaniel, Saar, and Titman 2005, Jones and Lipson 2003)
  - Plexus (Barber and Odean 2005, Jones and Lipson 2001)
  - Discount brokerage data on individuals (Odean 1998, 1999, Barber and Odean 2000, 2001, 2005)
  - Finnish data (Grinblatt and Keloharju 2000a, b)
  - Currency trading data (Evans and Lyons 2005, Froot and Ramadorai 2005b)

Existing Studies on Institutional Trading

- Proprietary data tend to have limited cross-sectional and time-series coverage
- Alternative is to use TAQ transactions data:
  - Classify buys and sells using Lee-Ready (1991) algorithm
  - Classify institutional trades using cutoff rule (Lee-Radhakrishna 2000)
  - Bennett-Sias (2001), Kim (2001) take this approach
- Our idea:
  - combine Spectrum with TAQ
  - find the best way to infer high-frequency institutional behavior from the transactions recorded in TAQ

Methodology

- **Goal:** To predict overall changes in institutional ownership from the general patterns of trading on a particular day.
- Find the best function mapping trade size to institutional behavior. This is distinct from a trade classification system.
- Regress changes in Spectrum ownership on cumulative quarterly trades of different sizes – from the ‘tape’.
- Regression function can be used to track institutional trading on a daily or intra-daily basis.
- How does institutional trading vary with characteristics of the stock (size, institutional ownership) and general market conditions (returns, volume, volatility)?

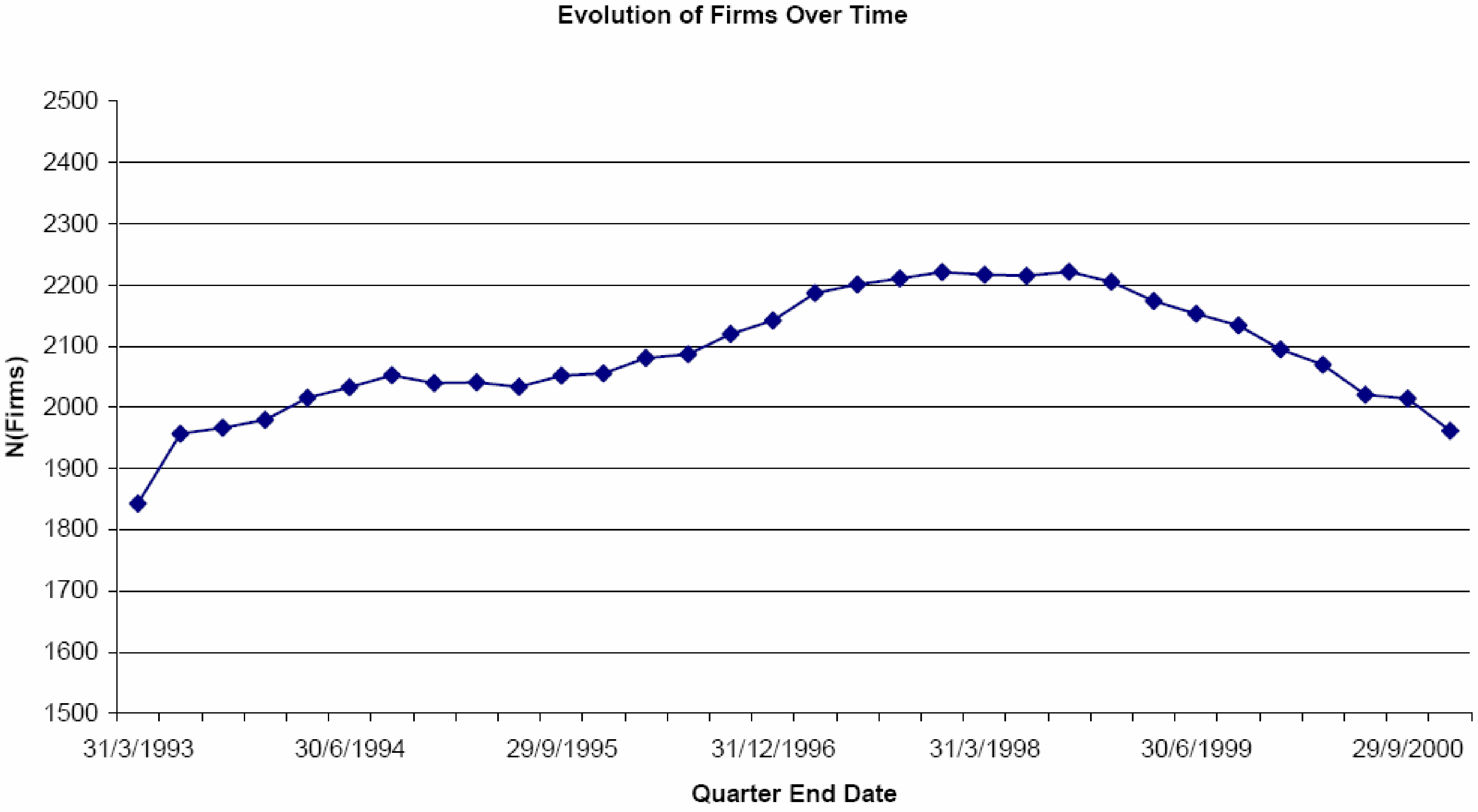
Data

- Institutional Ownership: (SPECTRUM)
  - 13-F filings of U.S. institutional investors. Required by SEC.
  - Institutions managing > US$ 1MM, in positions > 10,000 shares or US$ 200K, report quarterly holdings.

- 'Tape': (Transactions and Quotes (TAQ))
  - All trades and quotes on NYSE/AMEX since 1993. Sample ends December 2000.
  - Buy-Sell classification using Lee and Ready [1991] algorithm. 'Active' side of trade.
  - Trade size 'bins' using dollar cutoffs (0-2000, 2000-3000,…,900,000-1MM,>1MM).
  - Summed to quarterly frequency.

- CRSP:
  - Shares outstanding, permanent company numbers, market capitalization.

Summary Statistics



Summary Statistics: Annualized Percent of Firm Size

	All	Small	Q2	Q3	Q4	Large
Mean						
TAQ Total Buys	31.83	20.88	27.42	33.56	39.20	38.04
TAQ Total Sells	30.99	23.13	28.58	33.00	36.45	33.75
TAQ Unclassifiable	15.39	9.66	13.21	15.94	18.85	19.29
TAQ Total Volume	78.31	53.82	69.28	82.62	94.61	91.14
TAQ Net Imbalance	0.96	-2.13	-1.08	0.63	2.87	4.49
Spectrum Change	0.60	-1.31	0.29	1.73	1.49	0.77
Median						
TAQ Total Buys	23.84	13.72	18.76	24.85	31.40	30.58
TAQ Total Sells	23.90	15.80	20.63	25.58	29.90	27.41
TAQ Unclassifiable	11.57	5.70	8.73	11.47	15.04	15.81
TAQ Total Volume	60.42	36.43	49.26	63.03	77.00	74.35
TAQ Net Imbalance	0.55	-1.22	-0.62	0.15	1.62	3.09
Spectrum Change	0.43	-0.03	0.41	1.65	1.35	0.98
Standard Deviation						
TAQ Total Buys	13.48	11.00	13.04	13.84	14.23	12.82
TAQ Total Sells	12.40	11.32	12.55	12.81	12.83	11.30
TAQ Unclassifiable	6.79	5.69	6.69	7.00	7.03	6.21
TAQ Total Volume	31.68	27.09	31.27	32.62	33.06	29.46
TAQ Net Imbalance	5.07	4.87	5.07	5.22	5.15	4.23
Spectrum Change	8.94	7.48	9.40	9.87	9.59	8.03

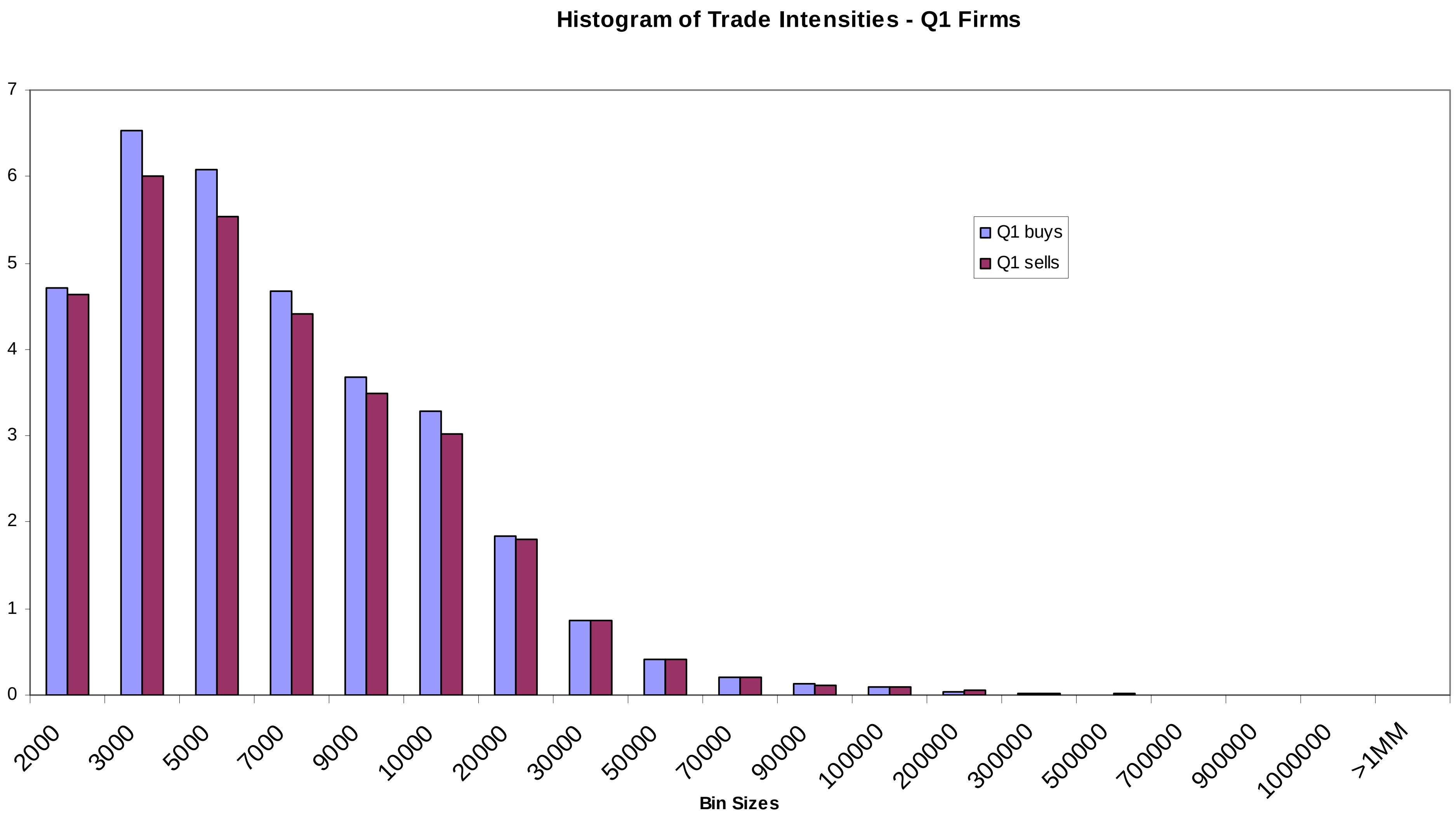
Summary Statistics: Annualized Percent of Firm Size

	All	Small	Q2	Q3	Q4	Large
Mean						
TAQ Total Buys	31.83	20.88	27.42	33.56	39.20	38.04
TAQ Total Sells	30.99	23.13	28.58	33.00	36.45	33.75
TAQ Unclassifiable	15.39	9.66	13.21	15.94	18.85	19.29
TAQ Total Volume	78.31	53.82	69.28	82.62	94.61	91.14
TAQ Net Imbalance	0.96	-2.13	-1.08	0.63	2.87	4.49
Spectrum Change	0.60	-1.31	0.29	1.73	1.49	0.77
Median						
TAQ Total Buys	23.84	13.72	18.76	24.85	31.40	30.58
TAQ Total Sells	23.90	15.80	20.63	25.58	29.90	27.41
TAQ Unclassifiable	11.57	5.70	8.73	11.47	15.04	15.81
TAQ Total Volume	60.42	36.43	49.26	63.03	77.00	74.35
TAQ Net Imbalance	0.55	-1.22	-0.62	0.15	1.62	3.09
Spectrum Change	0.43	-0.03	0.41	1.65	1.35	0.98
Standard Deviation						
TAQ Total Buys	13.48	11.00	13.04	13.84	14.23	12.82
TAQ Total Sells	12.40	11.32	12.55	12.81	12.83	11.30
TAQ Unclassifiable	6.79	5.69	6.69	7.00	7.03	6.21
TAQ Total Volume	31.68	27.09	31.27	32.62	33.06	29.46
TAQ Net Imbalance	5.07	4.87	5.07	5.22	5.15	4.23
Spectrum Change	8.94	7.48	9.40	9.87	9.59	8.03

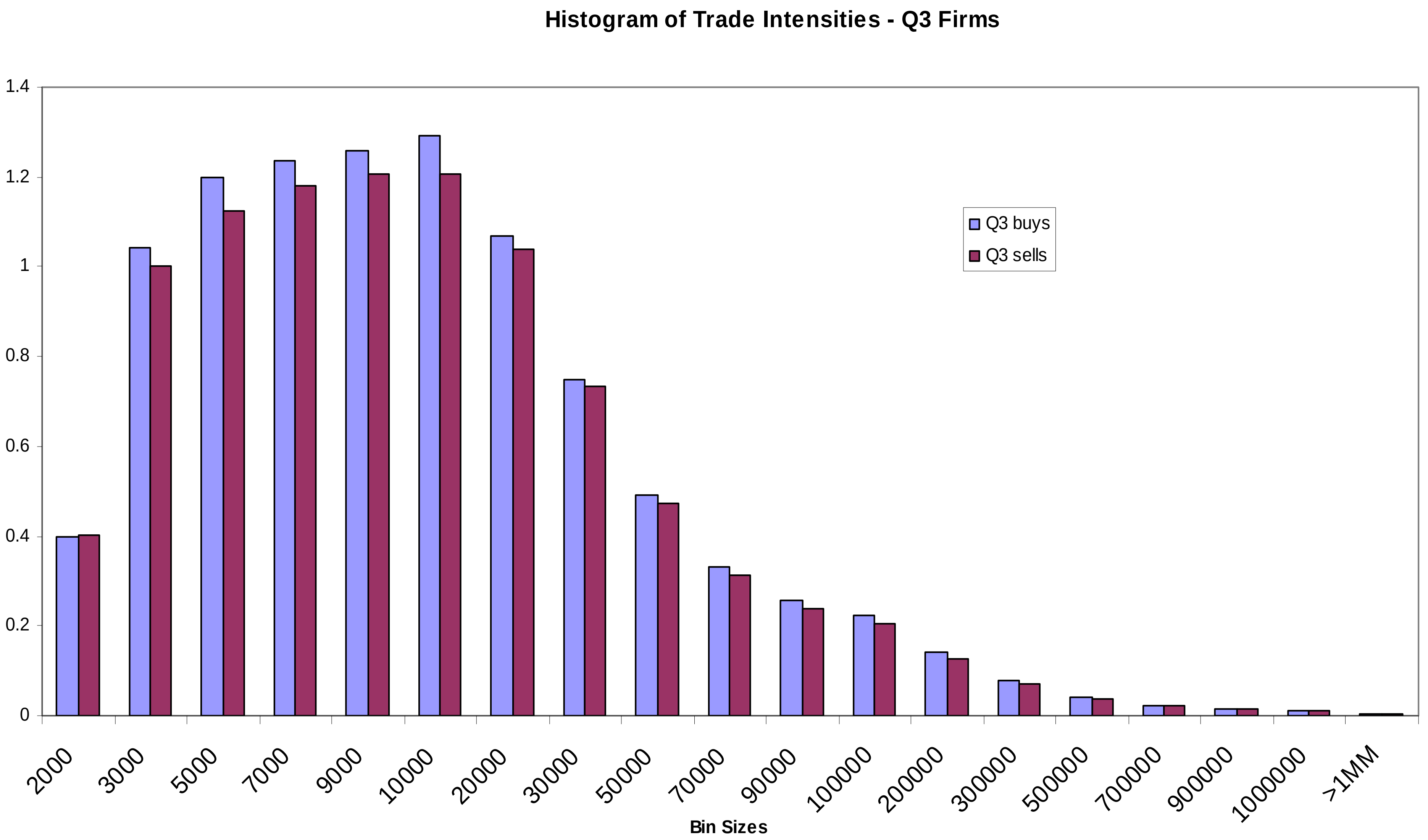
Summary Statistics: Annualized Percent of Firm Size

	All	Small	Q2	Q3	Q4	Large
Mean						
TAQ Total Buys	31.83	20.88	27.42	33.56	39.20	38.04
TAQ Total Sells	30.99	23.13	28.58	33.00	36.45	33.75
TAQ Unclassifiable	15.39	9.66	13.21	15.94	18.85	19.29
TAQ Total Volume	78.31	53.82	69.28	82.62	94.61	91.14
TAQ Net Imbalance	0.96	-2.13	-1.08	0.63	2.87	4.49
Spectrum Change	0.60	-1.31	0.29	1.73	1.49	0.77
Median						
TAQ Total Buys	23.84	13.72	18.76	24.85	31.40	30.58
TAQ Total Sells	23.90	15.80	20.63	25.58	29.90	27.41
TAQ Unclassifiable	11.57	5.70	8.73	11.47	15.04	15.81
TAQ Total Volume	60.42	36.43	49.26	63.03	77.00	74.35
TAQ Net Imbalance	0.55	-1.22	-0.62	0.15	1.62	3.09
Spectrum Change	0.43	-0.03	0.41	1.65	1.35	0.98
Standard Deviation						
TAQ Total Buys	13.48	11.00	13.04	13.84	14.23	12.82
TAQ Total Sells	12.40	11.32	12.55	12.81	12.83	11.30
TAQ Unclassifiable	6.79	5.69	6.69	7.00	7.03	6.21
TAQ Total Volume	31.68	27.09	31.27	32.62	33.06	29.46
TAQ Net Imbalance	5.07	4.87	5.07	5.22	5.15	4.23
Spectrum Change	8.94	7.48	9.40	9.87	9.59	8.03

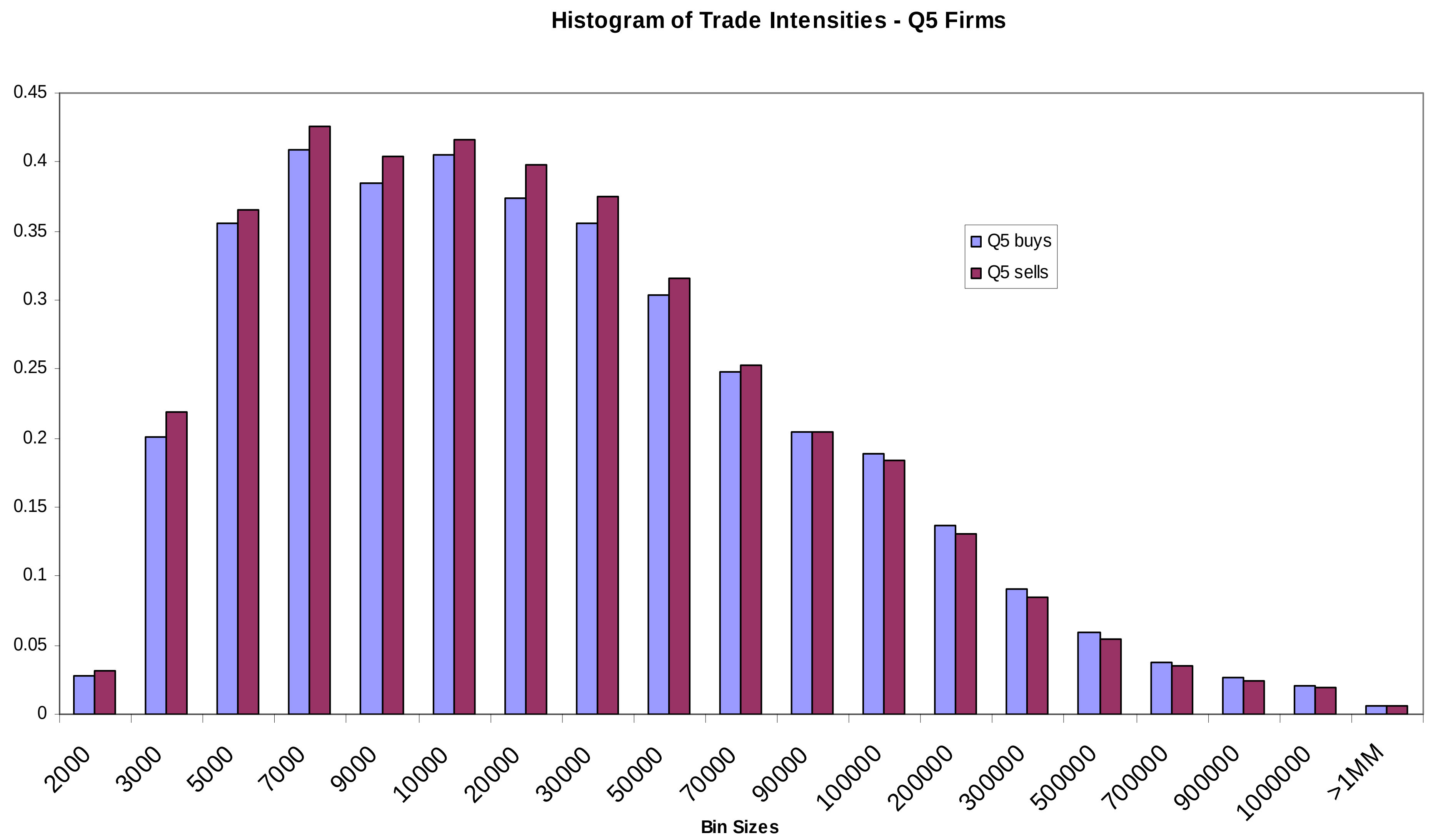
Histogram of Trade Intensities – Small Firms



Histogram of Trade Intensities – Mid Cap Firms



Histogram of Trade Intensities – Large Firms



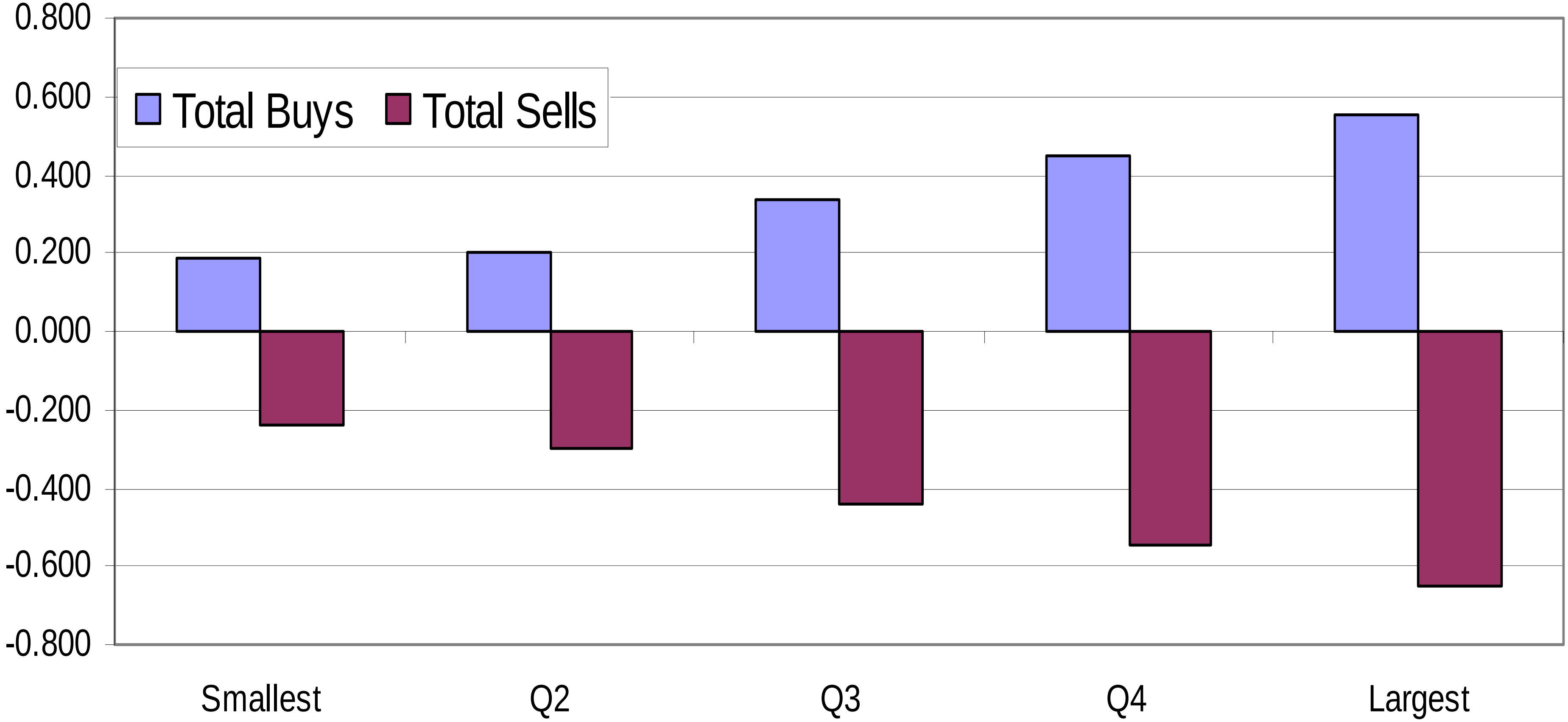
Regression Specifications – Total Flows

$$\underbrace{\Delta Y_{it}} = \alpha + \underbrace{\phi Y_{i,t-1}}_{\text{Lagged Level}} + \underbrace{\beta_U U_{it}}_{\text{Unclassifiable Volume}} + \underbrace{\beta_B B_{it}}_{\text{Total Buys}} + \underbrace{\beta_S S_{it}}_{\text{Total Sells}} + \varepsilon_{it}.$$

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + \underbrace{\beta_F F_{it}}_{\text{Total Net Flows}} + \varepsilon_{it}.$$

$$F_{it} = B_{it} - S_{it}$$

Total Flow Coefficients Across Size Quintiles

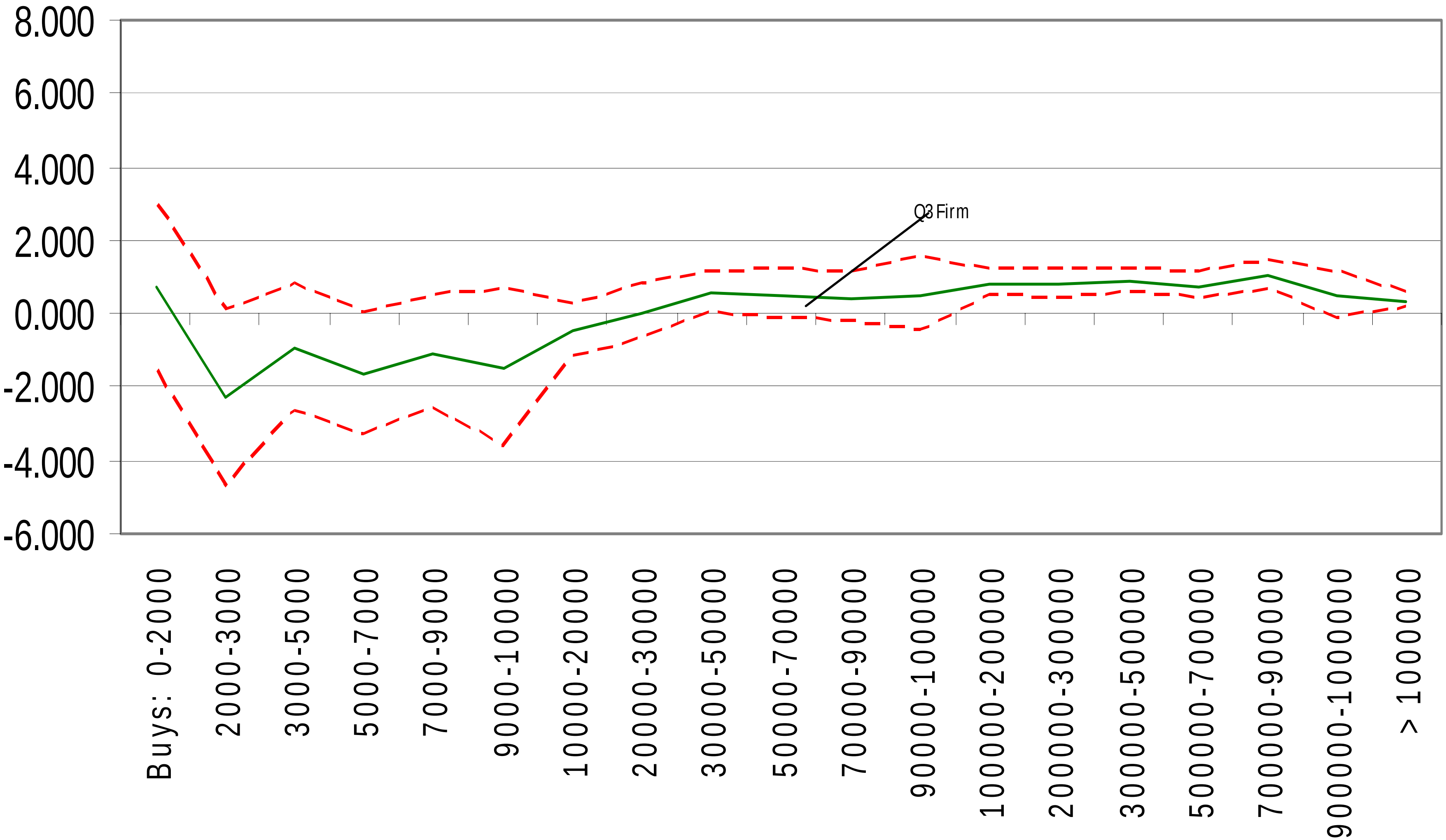


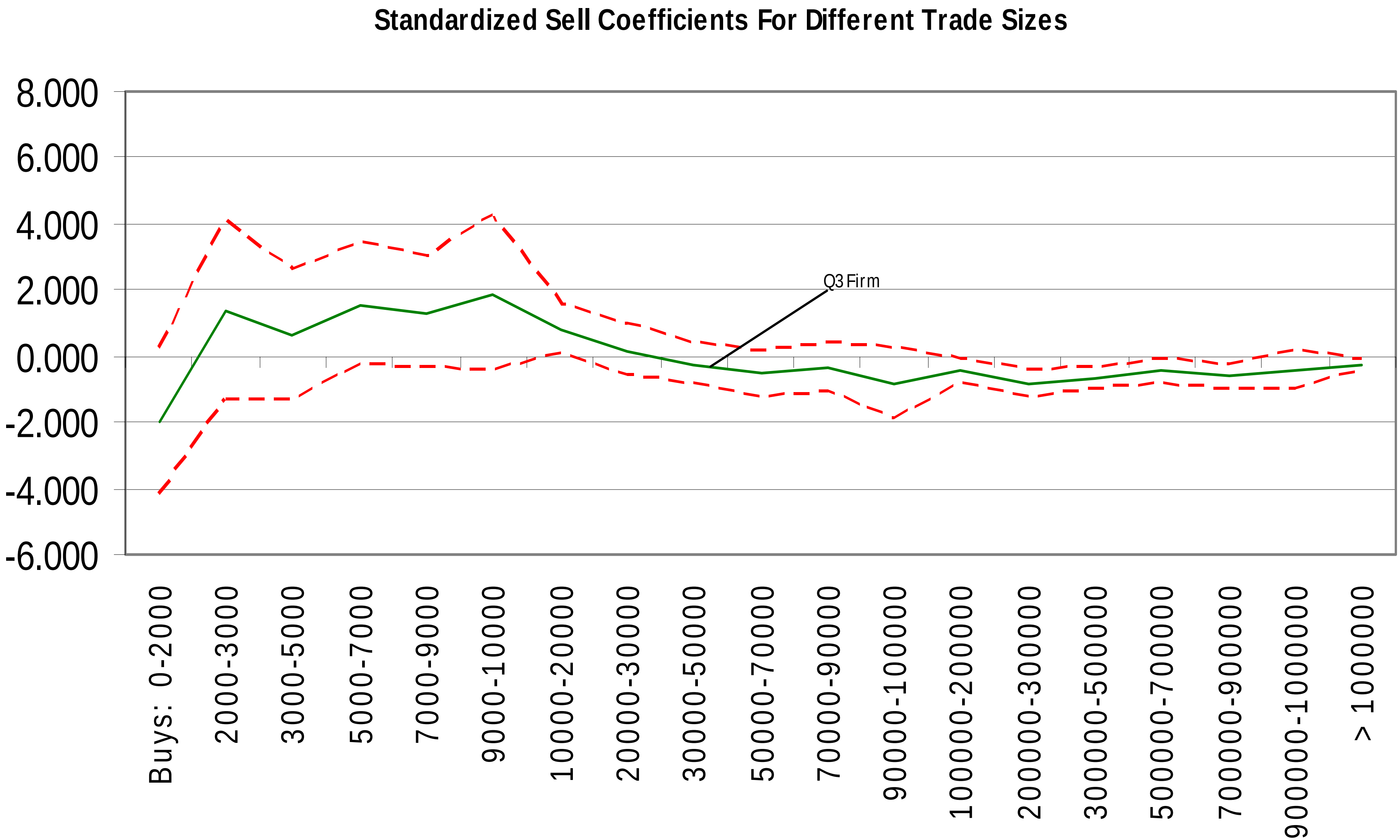
The Information in Trade Size

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + \underbrace{\sum_Z \beta_{BZ} B_{Zit}}_{\text{Bin Specific Buys}} + \underbrace{\sum_Z \beta_{SZ} S_{Zit}}_{\text{Bin Specific Sells}} + \varepsilon_{it}$$

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + \underbrace{\sum_Z \beta_{FZ} F_{Zit}}_{\text{Bin Specific Net Flows}} + \varepsilon_{it}$$

Standardized Buy Coefficients For Different Trade Sizes




Standardized Sell Coefficients For Different Trade Sizes
8.000
6.000
4.000
2.000
0.000
-2.000
-4.000
-6.000
Q3 Firm
Buys: 0-2000
2000-3000
3000-5000
5000-7000
7000-9000
9000-10000
10000-20000
20000-30000
30000-50000
50000-70000
70000-90000
90000-100000
100000-200000
200000-300000
300000-500000
500000-700000
700000-900000
900000-1000000
> 1000000

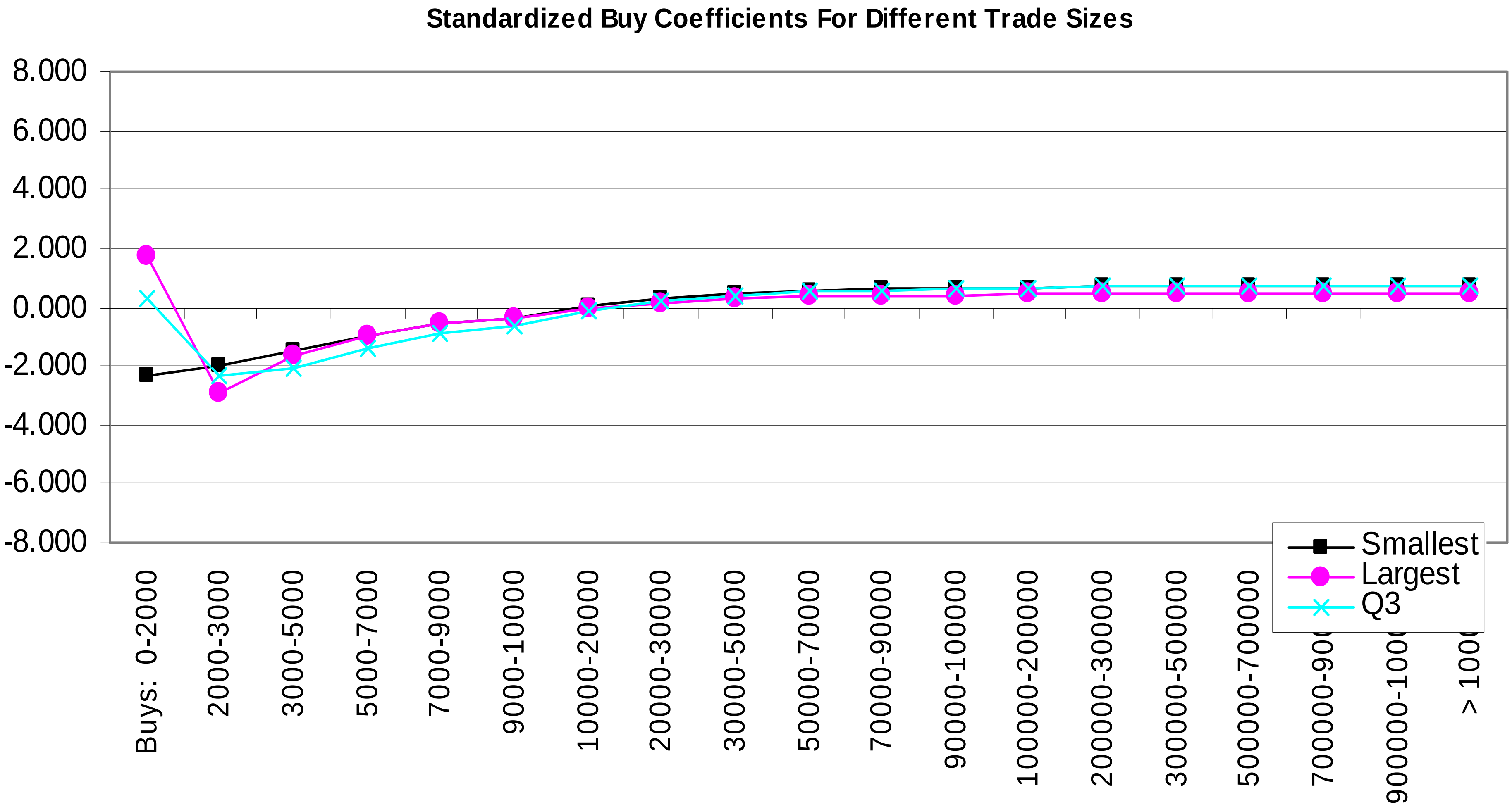
The Story So Far:

- Total buy (sell) volume predicts increasing (decreasing) institutional ownership
  - Institutions tend to buy at the ask and sell at the bid
  - Suggests that institutions demand liquidity rather than provide it.
- Buy volume in sizes between $2,000 and $30,000 is associated with decreasing institutional ownership
- Buy volume in larger sizes predicts increasing institutional ownership.
- Extremely small buys below $2,000 also predict increasing institutional ownership.
  - Scrum trades?
- All patterns are amplified as firm size increases (also as institutional ownership increases)

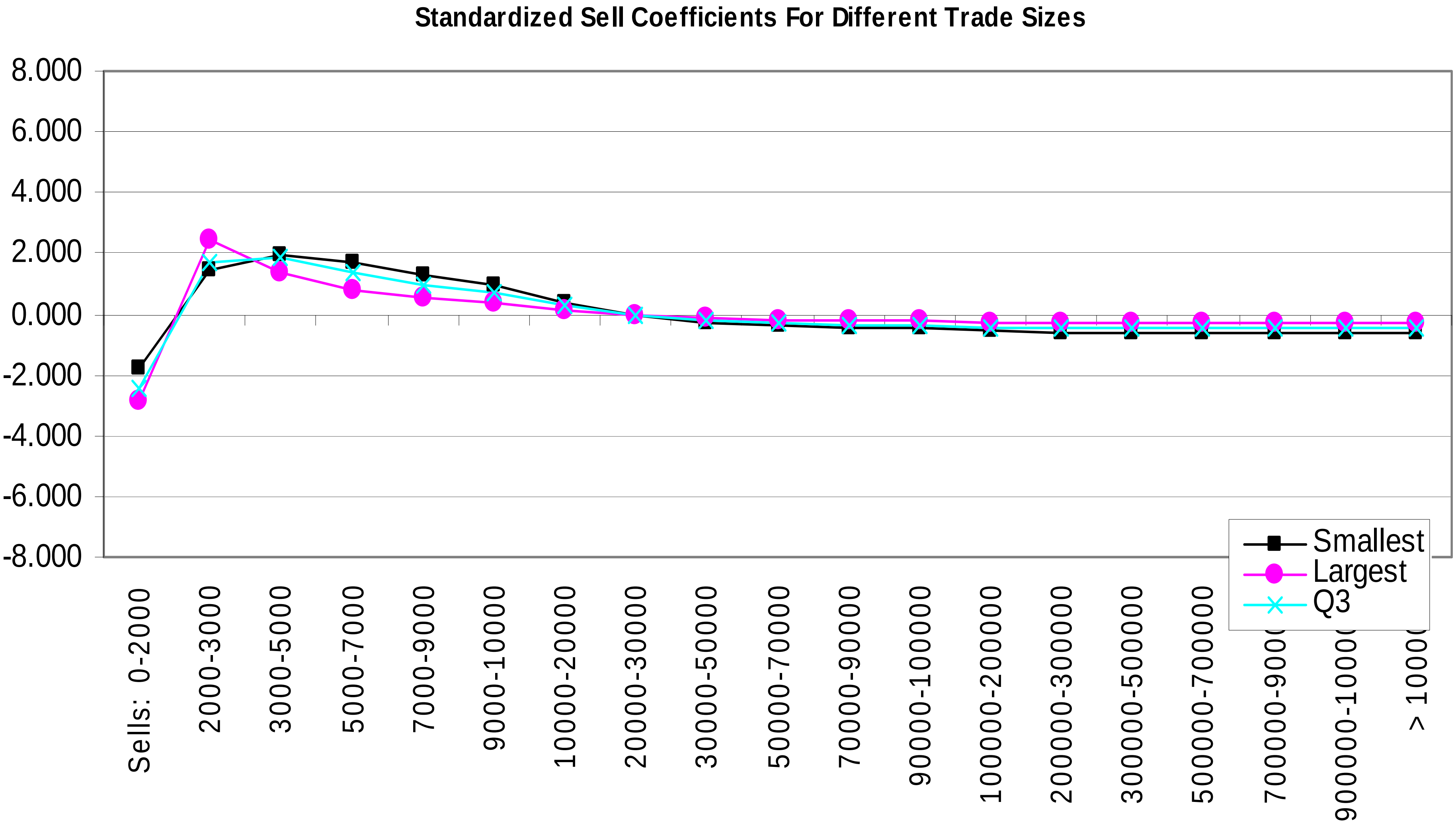
Smoothing the Effects of Trade Size

- Refine specification for a more accurate cutoff rule.
- Use Nelson and Siegel's [1987] nonlinear function (originally used for parsimonious yield-curve modeling).
- Reduces the number of parameters to estimate (weighting by bin midpoints).

$$\underbrace{\beta(Z)}_{\substack{\text{Coefficient for} \\ \text{bin Z}}} = b_0 + \underbrace{(b_1 + b_2)\left[1 - e^{-Z/\tau}\right]\frac{\tau}{Z} - b_2 e^{-Z/\tau}}_{\substack{\text{Negative exponential function: can capture a} \\ \text{range of shapes}}}$$



Standardized Buy Coefficients For Different Trade Sizes
8.000
6.000
4.000
2.000
0.000
-2.000
-4.000
-6.000
-8.000
Buys: 0-2000
2000-3000
3000-5000
5000-7000
7000-9000
9000-10000
10000-20000
20000-30000
30000-50000
50000-70000
70000-90000
90000-100000
100000-200000
200000-300000
300000-500000
500000-700000
900000
Smallest
Largest
Q3



Standardized Sell Coefficients For Different Trade Sizes
8.000
6.000
4.000
2.000
0.000
-2.000
-4.000
-6.000
-8.000
Sells: 0-2000
2000-3000
3000-5000
5000-7000
7000-9000
9000-10000
10000-20000
20000-30000
30000-50000
50000-70000
70000-90000
90000-100000
100000-200000
200000-300000
300000-500000
500000-700000
Smallest
Largest
Q3

But How Good is the Method?

R^2 **Matched Pairs of Cutoffs**	**Small**	**Q2**	**Q3**	**Q4**	**Large**
Lower: 2,000, Upper: 5,000	-0.080	-0.086	-0.033	0.010	0.068
Lower: 3,000, Upper: 10,000	-0.049	-0.062	-0.011	0.021	0.075
Lower: 3,000, Upper: 20,000	-0.021	-0.037	0.010	0.033	0.083
Lower: 3,000, Upper: 50,000	0.005	-0.014	0.032	0.046	0.091
Lower: 5,000, Upper: 100,000	0.004	-0.002	0.048	0.054	0.098
CRV Out of Sample					
Q1 1997: Q4 2000	0.073	0.061	0.072	0.080	0.116
Rolling One Period Ahead	0.077	0.072	0.092	0.089	0.117
N	12516	12609	12621	12732	12925
N(Firms)	1131	1357	1319	1161	735

Using the Method

- Current research (ongoing revision of the paper) constructs implied daily institutional flows and relates them to returns
- Estimate a VAR, with coefficient restrictions for parsimony, to characterize the joint dynamics of institutional net flows and returns
- Main results:
  - Net flows are positively autocorrelated
  - Returns predict net flows positively
  - Returns are negatively autocorrelated, particularly in small stocks
  - Net flows predict returns negatively in larger stocks (Q2-5)
  - The pattern in larger stocks is driven by sales rather than purchases: stocks that institutions purchase have a weak tendency to go up, but stocks that institutions sell have a strong tendency to rebound
  - Suggests that institutions take liquidity when they sell

The Basic VAR System

$$\begin{pmatrix} f_{i,t} \\ r_{i,t} \end{pmatrix} = \begin{pmatrix} \alpha_f \\ \alpha_f \end{pmatrix} + \begin{pmatrix} \phi_f & \phi_r \\ \rho_f & \rho_r \end{pmatrix} \begin{pmatrix} (1-\chi_f)\sum_{j=1}^{t-1} \chi_f^j f_{i,t-j} & (1-\lambda_f)\sum_{j=1}^{t-1} \lambda_f^j f_{i,t-j} \\ (1-\chi_r)\sum_{j=1}^{t-1} \chi_r^j r_{i,t-j} & (1-\lambda_r)\sum_{j=1}^{t-1} \lambda_r^j r_{i,t-j} \end{pmatrix} + \begin{pmatrix} u_{i,t}^f \\ u_{i,t}^r \end{pmatrix}$$

Basic VAR Results

	Small	Q2	Q3	Q4	Large	All
	$f_{i,t}$	$f_{i,t}$	$f_{i,t}$	$f_{i,t}$	$f_{i,t}$	$f_{i,t}$
χ_f	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
ϕ_f	0.664	0.636	0.633	0.681	0.691	0.671
	101.820	*108.652*	*121.177*	*153.230*	*170.689*	*294.700*
χ_r	**0.75**	**0.90**	**0.75**	**0.50**	**0.50**	**0.75**
ϕ_r	0.0007	0.0050	0.0060	0.0048	0.0056	0.0033
	18.276	*23.256*	*35.772*	*42.009*	*64.667*	*65.482*
R^2	0.0691	0.0629	0.0687	0.0872	0.1085	0.0782
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
λ_f	**0.98**	**0.00**	**0.00**	**0.90**	**0.90**	**0.90**
ρ_f	1.420	-0.836	-0.379	-0.493	-0.736	-0.472
	1.671	*-8.260*	*-6.740*	*-3.900*	*-5.288*	*-5.117*
λ_r	**0.50**	**0.75**	**0.00**	**0.75**	**0.75**	**0.50**
ρ_r	-0.322	-0.076	0.018	-0.035	-0.051	-0.155
	-28.938	*-6.390*	*5.635*	*-5.179*	*-7.945*	*-30.994*
R^2	0.0284	0.0010	0.0004	0.0002	0.0005	0.0073
N	549,979	578,707	592,630	621,557	668,413	3,015,050

Asymmetric Return Equation

A refinement allows for different effects of positive and negative net flows:

$$r_{i,t} = \alpha_r + \rho_f^b \left. \left(\left(1-\lambda_f\right) \sum_{j=1}^{t-1} \lambda_f^j f_{i,t-j} \right) \right|_{>0} + \rho_f^s \left. \left(\left(1-\lambda_f\right) \sum_{j=1}^{t-1} \lambda_f^j f_{i,t-j} \right) \right|_{<0} + \rho_r \left(1-\lambda_r\right) \sum_{j=1}^{t-1} \lambda_r^j r_{i,t-j} + u_{i,t}^r$$

Asymmetric Return Results

	Small	Q2	Q3	Q4	Large	All
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
λ_f	**0.98**	**0.00**	**0.00**	**0.90**	**0.90**	**0.90**
ρ_f^b	-2.344	1.238	0.498	0.285	0.795	1.363
	-1.447	*7.755*	*5.111*	*1.174*	*3.405*	*6.504*
ρ_f^s	-3.547	2.297	0.945	1.065	2.184	3.282
	-3.071	*14.981*	*11.698*	*5.564*	*9.276*	*4.346*
λ_r	**0.50**	**0.75**	**0.00**	**0.75**	**0.75**	**0.50**
ρ_r	-0.322	-0.076	0.017	-0.036	-0.052	-0.156
	-28.937	*-6.470*	*5.436*	*-5.308*	*-8.143*	*-31.067*
R^2	0.0284	0.0017	0.0007	0.0003	0.0007	0.0073
N	549,979	578,707	592,630	621,557	668,413	3,015,050

Interaction with Market Conditions

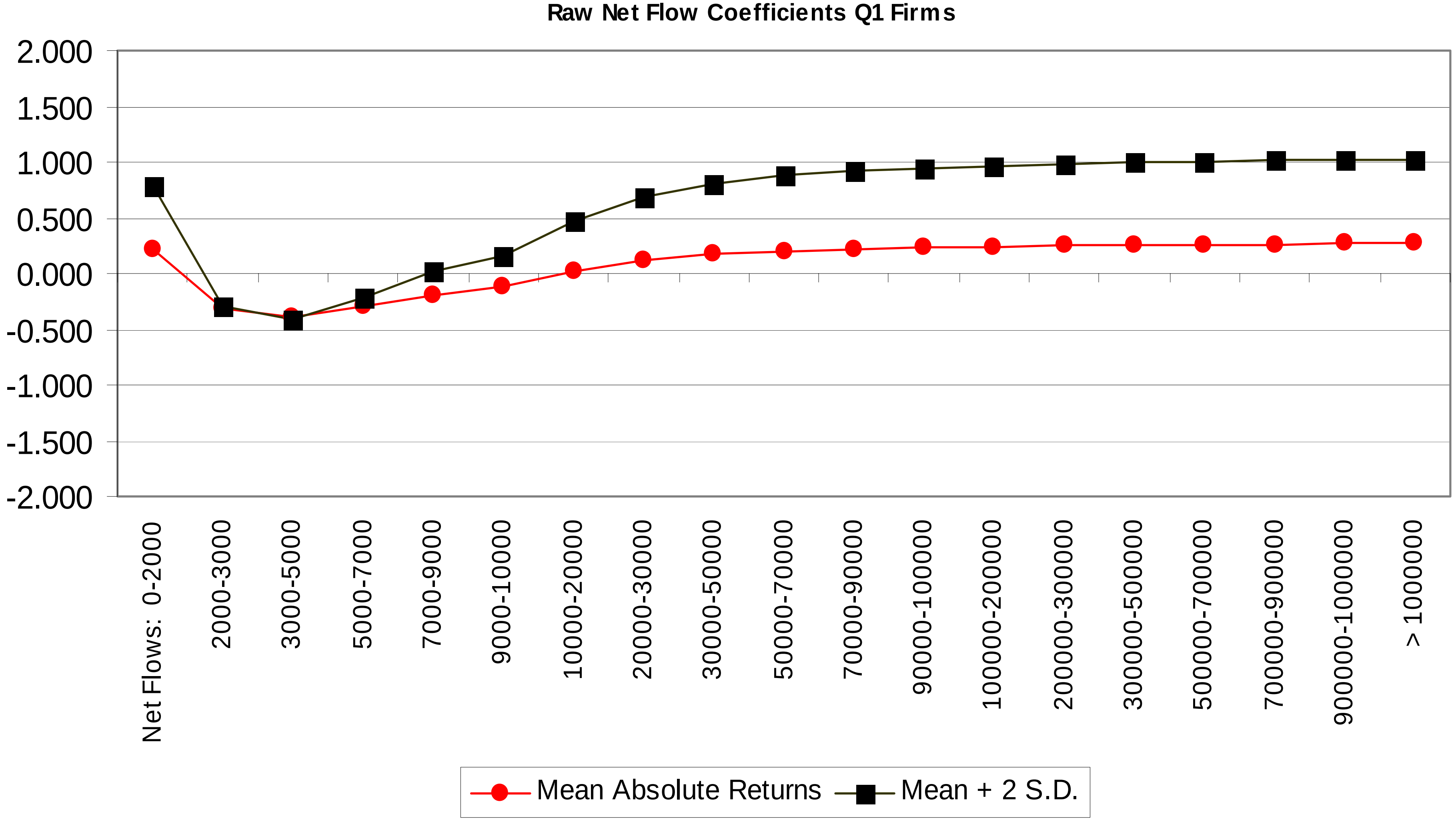
- How does institutional trading vary with movements in returns, volume, volatility and depth?
- Could we improve our method by re-specifying bins in different units (e.g. percent of daily volume or quoted depth)?
- We consider one interaction variable at a time.
- Nelson-Siegel parameters at the daily level are modeled as a linear function of the interaction variable.

$$\begin{aligned}\beta(Z, v_{id}) = {} & b_{01} + b_{02} v_{id} \\ & + (b_{11} + b_{12} v_{id} + b_{21} + b_{22} v_{id})[1 - e^{-Z/\tau}]\frac{\tau}{Z} \\ & - (b_{21} + b_{22} v_{id}) e^{-Z/\tau}\end{aligned}$$

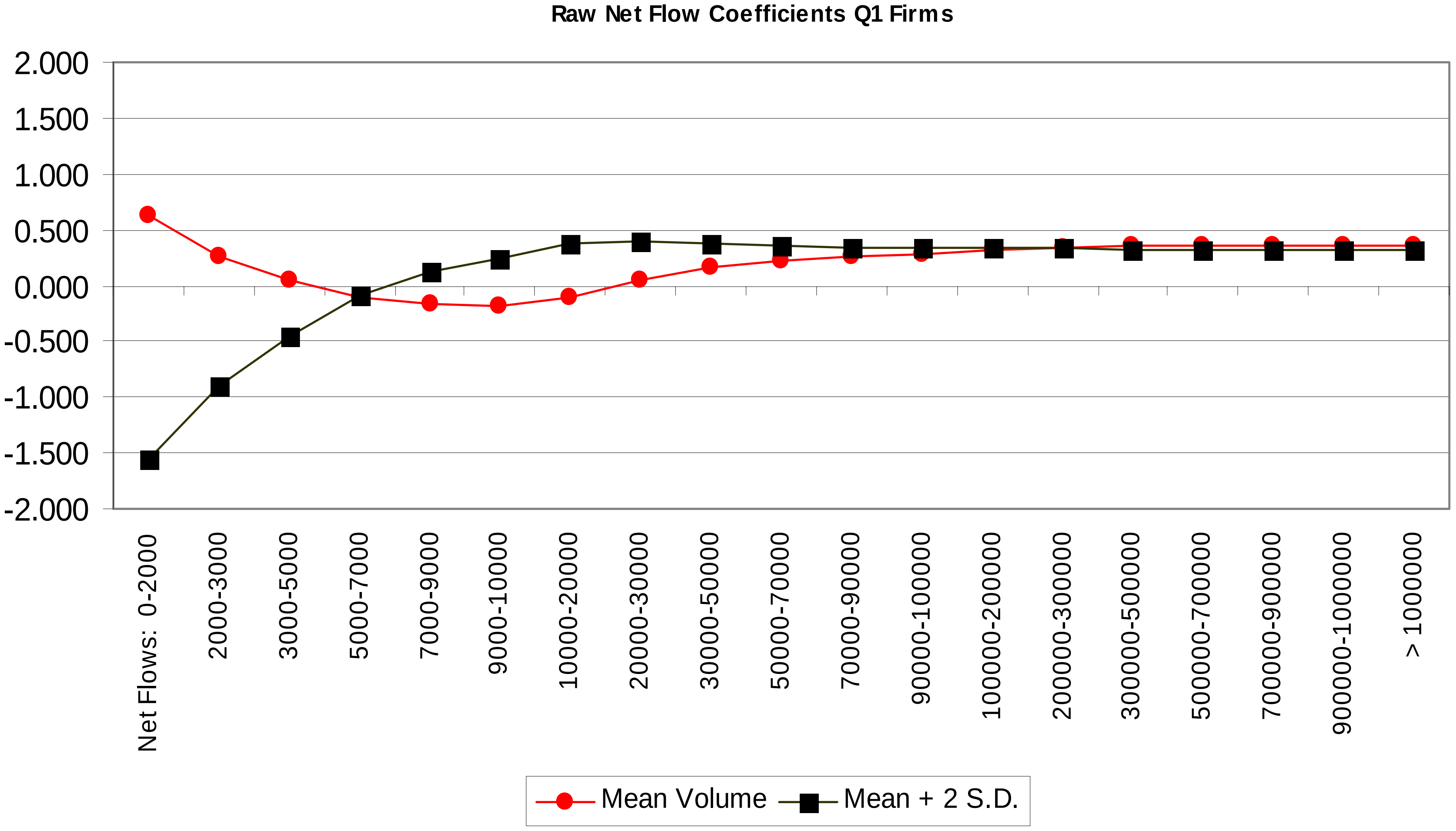
Interaction with Market Conditions

- Aggregation of the daily function to the quarterly level requires the assumption that unexplained net flows are uncorrelated with daily leads and lags of the regressors.
- More plausible for interactions that change the shape of the Nelson-Siegel function rather than its level.
- Less plausible for return interaction.

Volatility

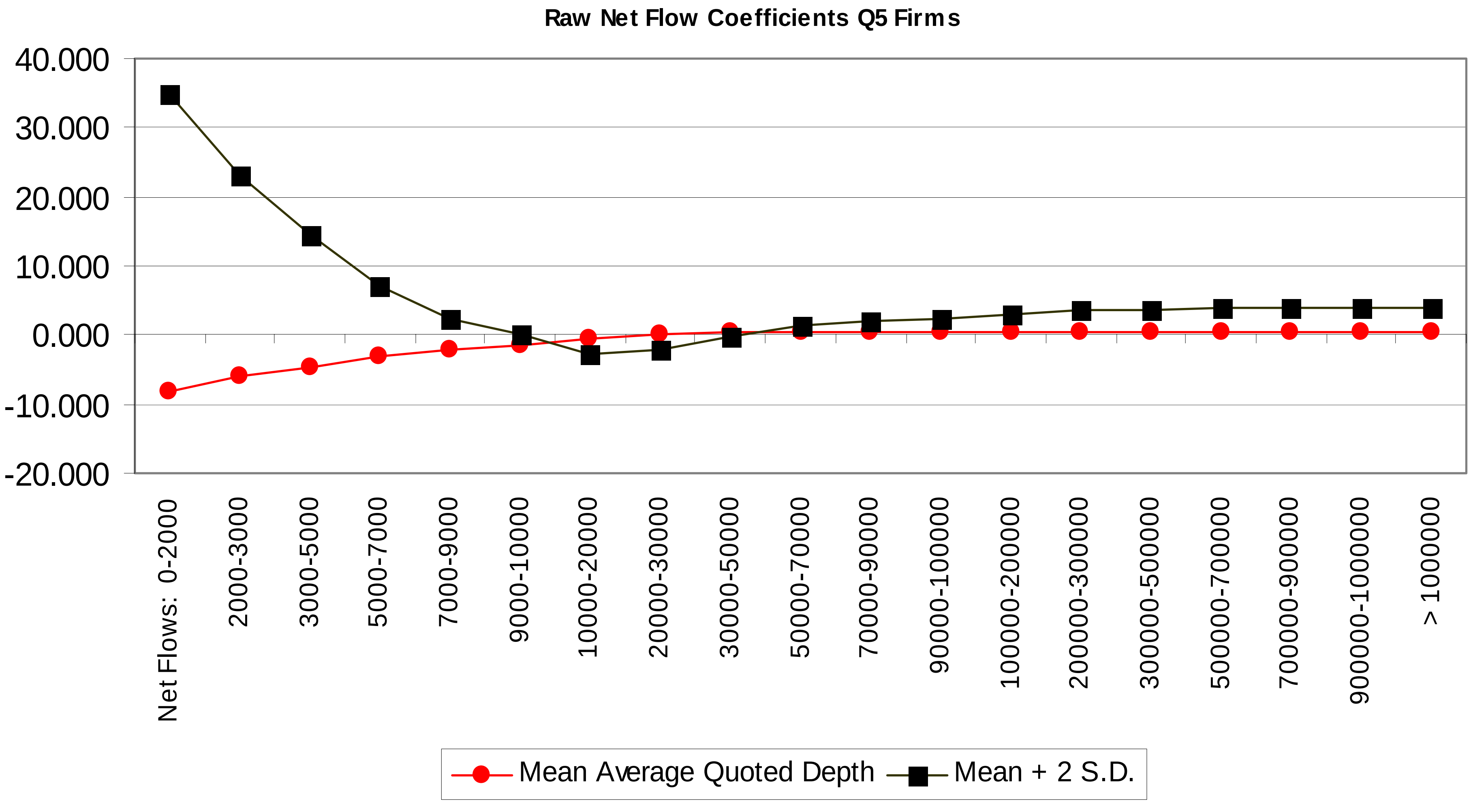

Raw Net Flow Coefficients Q1 Firms
2.000
1.500
1.000
0.500
0.000
-0.500
-1.000
-1.500
-2.000
Net Flows: 0-2000
2000-3000
3000-5000
5000-7000
7000-9000
9000-10000
10000-20000
20000-30000
30000-50000
50000-70000
70000-90000
90000-100000
100000-200000
200000-300000
300000-500000
500000-700000
700000-900000
900000-1000000
> 1000000
Mean Absolute Returns
Mean + 2 S.D.

Volume



Raw Net Flow Coefficients Q1 Firms
2.000
1.500
1.000
0.500
0.000
-0.500
-1.000
-1.500
-2.000
Net Flows: 0-2000
2000-3000
3000-5000
5000-7000
7000-9000
9000-10000
10000-20000
20000-30000
30000-50000
50000-70000
70000-90000
90000-100000
100000-200000
200000-300000
300000-500000
500000-700000
700000-900000
900000-1000000
> 1000000
Mean Volume
Mean + 2 S.D.

Depth



Raw Net Flow Coefficients Q5 Firms
40.000
30.000
20.000
10.000
0.000
-10.000
-20.000
Net Flows: 0-2000
2000-3000
3000-5000
5000-7000
7000-9000
9000-10000
10000-20000
20000-30000
30000-50000
50000-70000
70000-90000
90000-100000
100000-200000
200000-300000
300000-500000
500000-700000
700000-900000
900000-1000000
> 1000000
Mean Average Quoted Depth
Mean + 2 S.D.

Interaction With Market Conditions

- High volatility increases institutional trading intensity in the large bins: high demand for liquidity?
- High volume increases institutional trading in the medium-size bins: greater opportunity for stealth trading?
- High depth increases institutional trading in the small bins: greater opportunity to exploit the market-maker’s quotes?

Next Steps and Current Research

- We are currently investigating the relationship between daily TAQ flows generated using our method and daily stock returns.
- Our next project will explore whether institutional trading is different around earnings announcements, using the Nelson-Siegel function.
- We can measure institutional trading around earnings announcements and corporate actions. Do institutions 'arbitrage' individual underreaction to cashflow news?
- We can use the method to investigate institutional investor responses to a wide variety of corporate actions.