

Does Liquidity Matter When Crisis Is Brewing? Equity Returns and Intraday Liquidity in DAX Futures Market

Nonna Y. Sorokina
Pennsylvania State University

David E. Booth
Kent State University

Inspired by Mishkin (2011), we study a sample of high-frequency Eurex trading data from the peak of economic expansion preceding the global financial crisis of 2007-2009. Trading volume measures and bid-ask spread in DAX futures market explain DAX equity returns. Advanced variable selection and multinomial model with survey analysis identify most important explanatory variables under specific market conditions. The index futures trading volume, along with bid-ask spread, identify extreme values of the underlying equity index returns. Further, the Markov Chain Monte Carlo (MCMC) model uncovers change points in the non-monotonic relationship between futures liquidity and the equity index returns, and identifies conditions leading to expected negative returns.

Keywords: liquidity, futures, DAX, trading volume, bid-ask spread, Markov Chain Monte Carlo change point

INTRODUCTION

Asset price volatility, particularly in the overheated market, have long attracted attention of researchers, practitioners, regulators, and even general public (Alessi and Detken, 2011; Bernanke and Gertler, 2000, Mishkin, 2011, Smith et al., 1988). In this research paper, we study a period of excess liquidity on the peak of the market ahead of one of the deepest downturns not seen since the times of Great Depression. The dataset availability is limited to nearly 11 months of 2007, before the profound “liquidity gap” formed in global financial markets. We develop a model that connects equity returns with liquidity measures in the respective derivatives market.

Merton (1973) paved the road for integration of liquidity measures into asset pricing. Daily liquidity measures are a well-documented factor affecting asset prices (Amihud and Mendelson 1986, Datar et al. 1998, Pastor and Stambaugh 2003, Acharaya and Pedersen 2005). At the same time, asset price discovery is deeply rooted in trading dynamics of respective derivatives’ market (MacKinlay and Ramaswamy 1986, Yadav and Pope 1990, Lim 1992, Hasbrouck 2003, Miller et al. 2004). In this research, we connect the two strands of literature and study intraday liquidity in the futures market as an explanatory component of the underlying equity asset price. We also investigate ability of derivative liquidity measures to predict extreme points in the underlying asset returns.

German equity index DAX provides a convenient laboratory for this research. DAX is the index of the 30 blue chip German stocks traded on the Frankfurt Stock Exchange. The index is maintained by Deutsche Börse AG. DAX futures (FDAX) are highly active index derivatives traded on Eurex; they are also available for trading in the U.S, thus FDAX is a highly liquid trading instrument, available internationally. We analyze the correlation between trading volume – based measures of DAX futures liquidity and bid-ask spread of the futures with underlying equity returns. There are several important results in our study that provide innovative modeling approach and uncover important caveats in the role of liquidity in asset pricing.

First, there is significant presence of outliers (highly influential observations) in our sample, and they are concentrated among the observations, surrounding majority of the extreme return points. Moreover, the extremes are predictable by both contemporaneous and lagged trading volume and bid-ask spread. Additional analysis demonstrates that outliers help identifying unobserved effects (in our case, bid-ask spread, omitted in the initial specifications). Second, advanced variable selection methods identify trading volume, sell-to-buy ratio and bid-ask spread as most important general explanatory factors. Such array of DAX futures liquidity measures, derived from the high-frequency historical order books of Eurex, explains up to 20% of the variation in daily DAX returns in our sample. Further, multinomial logit model with Linearization and Jackknife support, highlights factors that are specifically relevant for identification of positive (negative) extremes. Finally, we confirm that the relationship between trading volume and bid-ask spread of the futures and index returns are non-monotonic. Markov Chain Monte Carlo (MCMC) change point model identifies the threshold values of liquidity measures that differentiate between negative and positive returns.

In summary, this paper presents useful methodology of explaining and predicting changes in the German equity market as reflected in the major equity index, DAX. Our results offer an insight into market dynamics for the wide audience of readers. Identifiable signs of approaching changes in the market may trigger regulatory actions, adjustments in corporate governance, as well as modifications in investment and trading strategies. We extend contribution of Czauderna et al. (2015) in three aspects. First, we draw the link between liquidity and assets across the markets for different instruments (equity and derivatives); second, our study highlights additional important measures of liquidity; third, we introduce a number of empirical methods that we not previously utilized for modeling of the liquidity-asset relations.

The paper is further structured as follows. We present an overview of the relevant academic literature in the “Literature” section; description of the data sources, data processing and variable construction may be found in the “Data and Variables Construction” part of the paper; the methodology, empirical results, and discussion for each section of the analysis are provided in the “Futures Volume and Equity Returns”, the “Additional Measures of Futures Liquidity and Underlying Asset Returns”, the “Extremes in the Index Prices and Futures Liquidity” and the “Change Points in the Relationship of Futures Liquidity and Index Returns” parts. The conclusions are drawn in the last section of this paper.

LITERATURE

Bernanke and Gertler (2000) deemphasize importance of assets markets for monetary policy decision and offer to focus on inflation. However, the vision of the importance of developments in the financial system for macroeconomic conditions was augmented following the financial crisis of 2007-2009 (Mishkin, 2011). In the past two significant downturns, Great Recession and COVID-19 crisis, regulators targeted financial stability more directly and their support methods included massive market intervention, such as quantitative easing, since then. Consequently, understanding of asset prices’ volatility grew critically important in macroeconomic context.

We find ample evidence of strong correlations between liquidity and asset prices on one hand, and interdependency of equity and derivative markets on the other hand in the finance literature. Merton (1973) laid a groundwork for pricing liquidity component into assets. Amihud and Mendelson (1986) find a positive relationship between bid-ask spread of stock prices and their returns by analyzing NYSE stock returns over the period of 1961-1980. Datar et al. (1998) provide an alternative test with turnover rate

employed as a liquidity measure. They all confirm that stock market liquidity plays a significant role in explaining stock returns. Further, Amihud et al. (1997) study liquidity implications for asset pricing and find that improvements in the Tel Aviv Stock Exchange microstructure lead to increased liquidity of the affected stocks followed by a permanent significant stock price increase. Amihud (2002) develops a liquidity measure based on dollar trading volume and explores liquidity premium from yet another perspective. Marshal and Young (2003) provide evidence of liquidity's significance for asset pricing from the Australian stock market. Martinez et al. (2004) confirm a relationship between systematic liquidity and asset prices, evident from the Spanish equity market. Pastor and Stambaugh (2003) document significance of aggregate liquidity for sensitivity of asset returns to individual asset's liquidity fluctuations. Acharaya and Pedersen (2005) build upon previous research and extend capital assets pricing by introducing covariance of a security's own expected liquidity with market liquidity as an important measure. Liu (2006) confirms significance of liquidity risk for capital assets prices. His study includes previously developed measures of liquidity and his own innovative measure based on the number of non-trading days in a year for a stock.

We also see that link between the equity market and derivatives market is well documented in the finance literature. Just to mention a few, Xing et al. (2010) find significant connection between stock options and their underlying equity. Pan and Poteshman (2006) provide evidence that option trading volume predicts future stock prices. Chan et al. (2002) use net trading volume and find that the stock market predicts developments in the options market, but do not find evidence of a link in the opposite direction. Chakravarty et al. (2004) use trading volume as their liquidity measure and show that option market contributes significantly to stock market price discovery. Cremers and Weinbaum (2010) document significance of liquidity for return predictability in their option-stock model. Beyond the option-based research cited above, MacKinlay and Ramaswamy (1986), Yadav and Pope (1990), Lim (1992), Miller et al. (2004) and others explore relationship between market developments and pricing of index futures and underlying index itself, including S&P 500, Nikkei 225 and FTSE 100. Hasbrouck (2003) finds, based on high frequency trading analysis, that price discovery of S&P 500 and Nasdaq 100 occurs in futures market. Booth et al. (1999) study intraday price discovery process for DAX stock index, index futures, and options. They conclude that futures market is dominant in price discovery; it is least costly and highly liquid.

Our work is in spirit of Czauderna et. al (2015). They add an innovative liquidity measure based on DAX ETF and offer a vector autoregressive models to show interconnectedness between liquidity and asset pricing for DAX. We take the idea a step further and show that liquidity in futures market provides a solid indicator of the direction of asset returns. Furthermore, our paper offers a rich methodological framework and includes outlier analysis and helps extracting valuable information from extreme and unusual values, instead of just simply disregarding them; advanced variable selection allows selecting most relevant factors among correlated candidates and avoids mistakes imposed by multicollinearity; multinomial logit analysis with survey data methodology mitigates limitations of the sample construction and data availability. Finally, simulation-based Markov Chain Monte Carlo change point analysis extends our findings beyond empirical experiments confined to the observable datapoints.

DATA AND VARIABLES CONSTRUCTION

We obtain liquidity - related data on DAX futures for the period between January 3, 2007 and November 23, 2007 from the Eurex historical order book database. Eurex Exchange is a part of the Eurex holding that operates a number of major European financial services companies. It is a public company wholly owned by Deutsche Börse AG. The exchange is well known worldwide for its massive, liquid, easily accessible equity-based derivatives operations, thus the environment is ideal for our purpose. Eurex maintains a dataset of complete daily order books for each instrument with a record of each quote and trade including expiration month, buy/sell, price, beginning and ending time stamp precise to millisecondsg volume, as a total dollar amount of all trades, interval length, order size, number of orders, and aggressive/passive trade indicator.

We compute four measures of futures' liquidity using daily order book data: daily trading volume, as a total dollar amount of all trades throughout the day (Karpoff, 1987), and a more sophisticated volume-related variable, similar to the net trading volume of Chan et al. (2002), - a measure of selling pressure, represented by the ratio of aggressive sell to aggressive buy trades in a trading day. In addition to the volume-related variables, we compute average bid-ask spread (Amihud and Mendelson, 1986) for each day. For every trade, we find all orders that are open at that same time and compute spread; further we take average bid-ask spread of the records corresponding to all unique timestamps associated with trades throughout the day. We adapt the "non-trading days in a year", a variable developed by Liu (2006) for daily trading of individual stocks, to our high-frequency derivatives environment. In our analysis the variable measures number of non-trading minutes in each trading day. While trading volume and bid-ask spreads were widely used to test high-frequency data, the fourth measure was only tested in the datasets with daily frequency before. We show successful transition of the daily measure into the high-frequency environment.

TABLE 1
SUMMARY STATISTICS AND PEARSONS CORRELATIONS FOR THE VARIABLES

Panel I: Summary statistics				
	Mean	Median	Std	Number Obs
Volume	183675	170307	74134	230
AggrSellToBuy	0.9965	0.9972	0.0448	230
NonTradingMin	10.04	0	62.08	230
BidAskSpread	41.99	55.95	24.96	230
ReturnDAX	0.00066798	0.001184	0.00973	230

Panel I provides standard summary statistics: mean, median and standard deviation for dependent and independent variables used in this study: return on DAX; trading volume, aggressive sell to buy, number of non-trading minutes in a day, and bid-ask spread of DAX futures.

Panel II. Pearsons Correlations for DAX Futures Liquidity Measures

	Volume	AggrSellToBuy	NonTradingMin	BidAskSpread
Volume	1	0.3136 *** <i><.0001</i>	-0.2705 *** <i><.0001</i>	0.1731 <i>0.0085</i>
AggrSellToBuy	0.3136 *** <i><.0001</i>	1	-0.5823 *** <i><.0001</i>	0.0764 <i>0.2488</i>
NonTradingMin	-0.2705 *** <i><.0001</i>	-0.5823 *** <i><.0001</i>	1	-0.0300 <i>0.6513</i>
BidAskSpread	0.1731 *** <i>0.0085</i>	0.0764 <i>0.2488</i>	-0.0300 <i>0.6513</i>	1

The pairwise Pearson correlation coefficients are presented in this table along with respective p-values that reflect their statistical significance. Panel A depicts correlations between independent variables, panel B – correlations between closing values of the indices, used as dependent variables, and panel C – correlations between returns on the indices, also dependent variables.

On the equity side, daily adjusted closing values of DAX for the period January 03, 2007 – November 23, 2007 are obtained from Bloomberg. Returns on equity index are computed from closing index values as follows:

$$R_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \quad (1)$$

The summary statistics of the data are presented in Table 1, Panel I. We report all independent variables of interest for each of 230 days of futures data available to us. We also get 230 matching observations of DAX. The pairwise Pearson correlation coefficients for the independent variables are provided in Table 1, Panel II. The most significant is negative correlation between NonTradingMinutes and AggrSellToBuy. However, correlations between Volume and AggressiveSellToBuy (positive) and Volume and NonTradingMinutes (negative) may deserve some attention. The last statistically significant positive correlation between BidAskSpread and Volume is quite weak.

METHODOLOGY, RESULTS, AND DISCUSSION

Futures Volume and Equity Returns

Ample evidence in literature documents predictive power of trading volume for equity returns (Karpoff, 1987; Amihud et al., 1997; Amihud, 2002; Pastor and Strambaugh, 2003; Acharya and Pedersen, 2005; Liu, 2006; Czauderna et al., 2015). At the same time, a strong strand of research is dedicated to relationship between derivatives and equity market (Booth et al., 1999; Chan et al., 2002; Chakravarty et al., 2004; Xing et al., 2010). We begin our study with a simple model of relationship between trading volume in derivative market and returns on the underlying equity. The initial regression model includes trading volume – based liquidity measures of the DAX futures as independent variables and allows us to validate connectedness of futures liquidity and equity returns:

$$R_{DAX} = \beta_0 + \beta_1 Volume_{FDAX} + \beta_2 SellToBuy_{FDAX} + \epsilon \quad (2)$$

where R_{DAX} – return on DAX

$Volume_{FDAX}$ - trading volume of the DAX futures

$SellToBuy_{FDAX}$ – ratio of the aggressive sell trades and aggressive buy trades on DAX futures

The estimates show strong and statistically significant relationship between DAX returns and both liquidity measures, daily futures trading volume and ratio of aggressive sell trades to aggressive buy trades in a trading day as presented in Panel I of Table 2 (contemporaneous). All models based on lagged volume parameters (one, three, five and ten days) produced no statistically significant estimates in this specification. We will continue exploring predictability of DAX returns with lagged liquidity measures in the subsequent sections of the paper. Both volume and selling pressure measures have negative impact on stock returns, i.e. as futures' trading volume and/or selling pressure in futures increase, stock returns decrease.

We also run univariate versions of regression (2) according to the specifications (2a) and (2b) below on each of the four indices.

$$R_I = \beta_0 + \beta_1 Volume_{FDAX} + \epsilon \quad (2a)$$

$$R_I = \beta_0 + \beta_1 SellToBuy_{FDAX} + \epsilon \quad (2b)$$

where R_I – return on DAX

The results, presented in panel II of Table 2, confirm that both measures of DAX futures trading volume are strong explanatory variables for DAX returns.

The daily and more frequent financial data are well known for non-normal distributions and increased chance of occurrence of influential data points (outliers). Some view outliers as noisy and annoying data irregularities (Barnett and Lewis, 1978). However, outliers may carry valuable information and may also be responsible for famous anomalous effects. For example, Knez and Ready (1997) apply robust regression and find that the Fama and French (1992) size effect comes from the 1% most extreme observations in their sample. We analyze outliers in our DAX dataset, recognizing outliers by Cook's distance that was introduced by Cook (1977) for the purpose of identifying outliers in OLS regression. Cook's distance is the most commonly used estimate of influence of a data point in a least squares regression. It measures the effect of deleting a given observation and identifies both outliers and high leverage points. We choose a Cook's distance of $4/(n-k-1)$ as the cutoff for an outlier, where n = number of observations and k = number of independent variables, as suggested by Belsley et al. (2005). This methodology is similar to one employed by Sorokina et al. (2013) in an event study application.

TABLE 2
EQUITY INDICES AND FUTURES TRADING VOLUME MEASURES

Panel I: Bivariate regression			
	Volume		AggressiveSellToBuy
DAX, Contemporaneous	-0.000000047310900 *** <i><.0001</i>		-0.040220000000000 *** <i>0.0034</i>
Panel II: Univariate regressions, FDAX volume measures			
DAX (Underlying)	-0.0000000548967 *** <i><.0001</i>		-0.06327 *** <i><.0001</i>

This table reports the estimates of the regressions of DAX returns on two trading volume – based measures of DAX futures liquidity. Panel I shows the estimates of bivariate regression of DAX returns on trading volume and ratio of aggressive sell to aggressive buy trades. Panel II presents results of univariate regressions for the same independent variables.

In our study period, we find six extreme index value points, 3 significant maximums and 3 minimums. DAX market peaks occurred on 02/27/2007, 07/16/2007, and 10/31/2007; and market bottoms were on 11/19/2007, 09/10/2007 and 03/14/2007. It turns out that outliers surround five of the six extreme points in the index return trend. These results are detailed in Table 3. Moreover, there are only two occurrences of outliers in the whole sample that do not correspond to the extreme index value points. Those occurrences, however, coincide with brief, but significant, deeps of the index (06/06-06/15 and 08/09-08/16). The findings suggest strong association between outliers in the regression of DAX return on its futures liquidity measures and extreme return values. The liquidity measures of index futures explain a significant portion of underlying index returns well, on average. However, correlation between liquidity measures and returns significantly deviates from average at the return extreme points. The information uncovered by outliers suggests a possibility of identification of extreme points with liquidity measures. There is(are) also potential change point(s) in the regression prediction line of DAX return with liquidity measures as predictors.

Additional Measures of Futures Liquidity and Underlying Asset Returns

Intraday trading volume of the futures proves as a powerful predictor of index returns in our sample. However, trading volume does not tell a complete story. Starting with Amihud and Mendelson (1986) spread is shown theoretically and empirically to capture an important caveat in asset pricing. Spread represents illiquidity and measures willingness of an investor to transact with a concession rather than wait for more favorable environment. We augment our basic model with average intraday bid-ask spread of futures trades. Further, Liu (2006) highlights an interesting dimension of trading continuity. His measure of illiquidity counts gaps in trading as a number of non-trading days in the year. We adapt this variable to high-frequency environment and count number of non-trading minutes in a day for the fourth explanatory variable in our model. In the initial step of model development, we run univariate regressions of DAX returns with each of the four futures liquidity measures, as shown below.

**TABLE 3
OUTLIERS AND EXTREME DAX VALUE POINTS**

Trading Date	Cooks Distance	Outlier?	Extreme?
26-Feb-07	0.029957534	X	Peak
27-Feb-07	0.048989019	X	
28-Feb-07	0.023284405	X	
.....			
13-Mar-07	0.093864384	X	
14-Mar-07	0.099729229	X	Bottom
.....			
05-Jun-07	0.032597917	X	
06-Jun-07	0.019088655	X	
07-Jun-07	0.001469528		
08-Jun-07	0.021903298	X	
11-Jun-07	0.000454077		
12-Jun-07	4.14607E-05		
13-Jun-07	0.078552349	X	
14-Jun-07	0.027660974	X	
15-Jun-07	0.018420734	X	
.....			
11-Jul-07	0.019765832	X	
12-Jul-07	0.000352664		
13-Jul-07	2.30101E-05		
16-Jul-07	0.002709869		Peak
.....			
25-Jul-07	0.033230665	X	
.....			
09-Aug-07	0.018119938	X	
10-Aug-07	0.057244485	X	
13-Aug-07	0.001329785		
14-Aug-07	0.000542974		
15-Aug-07	0.024828477	X	
16-Aug-07	0.044601773	X	
.....			
10-Sep-07	0.002576105		Bottom
.....			
31-Oct-07	0.013687169	X*	Peak
.....			
19-Nov-07	0.023427364	X	Bottom

* the value is quite close to the critical threshold of 0.01762

The above presented list of observations includes dates surrounding Peaks and Bottoms of DAX returns trend. The majority of observations are classified as outliers in the regression of DAX returns on DAX futures liquidity parameters (Table 3, Panel I). Cooks distance is reported as well; it serves as a basis for outlier detection.

$$R_{DAX} = \beta_0 + \beta_1 X_{FDAX} + \epsilon \quad (3)$$

where X_{FDAX} is trading volume, aggressive sell to aggressive buy trades ratio, bid-ask spread or number of non-trading minutes.

TABLE 4
DAX FUTURES LIQUIDITY AND UNDERLYING INDEX RETURNS

Panel I: Four variable specifications

R2 (Adj)	Volume	SellToBuy	NonTradMin	BidAskSpread
0.1689	-0.0000000548967 *** <i><.0001</i>			
0.0806		-0.06327 *** <i><.0001</i>		
0.0125			0.00002035 ** <i>0.0494</i>	
0.0328				-0.00003347 *** <i>0.0034</i>
0.1941	-0.0000000471993 *** <i><.0001</i>	-0.03876 *** <i>0.0047</i>		
0.1984	-0.0000000485962 *** <i><.0001</i>	-0.05186 *** <i>0.0015</i>	-0.00001713 <i>0.1374</i>	
0.2084	-0.0000000459859 *** <i><.0001</i>	-0.05061 *** <i>0.0018</i>	-0.00001631 <i>0.1547</i>	-0.00002037 ** <i>0.0513</i>

Panel II: 4-variable model, additional diagnostics

White p-values	<i><.0001</i>	<i>.0004</i>	<i>0.0350</i>	<i>0.0167</i>
VIFs	<i>1.15343</i>	<i>1.57829</i>	<i>1.53497</i>	<i>1.03285</i>

Panel III: Lasso – optimal model

	-0.00000003649605	-0.042006	-0.000030156
t-value	-4.38	-3.08	-2.87

In Panel I the results of the OLS univariate and stepwise regression analysis are provided. Return on DAX index is regressed on trading volume, ratio of aggressive sell to aggressive buy trades, number of non-trading minutes, and bid-ask spread individually. Then all variables are added one by one into a combined regression model. In panel II additional diagnostics are provided for the four-variable model: heteroskedasticity-robust White p-values and variance inflation factors – an indicator of multicollinearity. In Panel III the results of the advanced variable selection model lasso are shown. The estimates are also robust to multicollinearity.

Further, we include all four predictors into a multivariable specification, in a stepwise selection process:

$$R_{DAX} = \beta_0 + \beta_1 Volume_{FDAX} + \beta_2 SellToBuy_{FDAX} + \beta_3 NonTradeMin_{FDAX} + \beta_4 BidAsk_{FDAX} + \epsilon \quad (4)$$

The results are presented in Table 4, panel I. We can see that each measure of futures liquidity helps explaining a part of the variation in the underlying index returns, when considered separately. Volume, SelltoBuy and BidAskSpread all have negative impact on returns, while NonTradingMinutes's isolated effect is positive and significant. However, the NonTradingMinutes coefficient changes sign and loses its statistical significance when combined with other liquidity measures. The greatest explanatory power is associated with trading volume and selling pressure, the adjusted R-squared for these two variables is 16.89% and 8.06% respectively, adjusted R-squared for the bid-ask spread reaches 3.28% and for the number of non-trading minutes it is 1.25%. Higher values of all variables, except for the number of non-trading minutes, are associated with lower returns. In the multivariable specifications, all but non-trading minutes hold their significance and sign; the explanatory power of the model grows with addition of every new variable, and adjusted R-squared reaches 20.84% for the four-variable model.

Some of the measures of liquidity in our study are quite strongly correlated, as shown in Panel II of Table 1 and we are concerned with a potential issue of multicollinearity in the multivariable model. In case of multicollinearity, a multivariable regression model is valid. However, the impact of an individual parameters may be misattributed and coefficients of individual predictors and their standard errors may not be reliable. We take some precautions to rule out the problem and apply the lasso method originally described by Tibshirani (1996). Traditionally, variance inflation factor (VIF) analysis helps demonstrating how much the coefficient increased as a result of collinearity and provides means for adequate estimation of the coefficients. Ridge regression of Hoerl and Kennard (1970) allows accurate estimation of coefficients in datasets subjected to collinearity. The researchers, though, are not only interested in accurate estimates, but also in selection of a model with minimal number of predictors and maximal explanatory power. The lasso combines benefits of variance inflation reduction provided by ridge regression with robust variable selection.

In addition to proper treatment of collinearity, lasso helps avoiding problems of stepwise selection. Tibshirani (1996), among many others, points out to the overall instability of estimates and especially poor out-of-sample performance of stepwise regressions. A selection of variables in stepwise method often does not provide the best set of explanatory variables according to Austin and Tu (2004).

The lasso results are presented in Panel III of Table 4 in the form of the optimal model selected by lasso. As we can see, the NonTradingMinutes variable is not included in the optimal suggested model, while the coefficients of other explanatory variables are estimated as negative and statistically significant. Interestingly, the set of variables selected by lasso matches the result of stepwise selection in our case. We confirm our model selection with more advanced adaptive lasso methods of Zou (2006) and Wang and Leng (2007). However, lasso reduces the estimated size of the influence of volume measures and increases size of the coefficient of bid-ask spread. The adjusted R-squared of the final model is 18.95% and explanatory power as measured by traditional R-squared is 20.01%.

Extremes in the Index Prices and Futures Liquidity

Positive and negative extremes in returns trend resolve anxiety in the market that makes conditions surrounding those culminative moments in price discovery important for academics, practitioners, and regulators. Outlier analysis, outlined in section 4.1 of this study, identifies connection between developments in the futures market and underlying index peaks and bottoms. In this section we show how to identify DAX extreme returns points with futures liquidity measures. The multinomial logit models help distinguish between positive extremes and negative extremes (peaks and bottoms), and non-extremes of the index. Our analysis begins with univariate multinomial logit regressions that include each of four liquidity measures.

$$Extereme_{DAX} = \beta_0 + \beta_1 X_{FDAX} + \epsilon \quad (5)$$

where X_{FDAX} takes a value of volume, sell-to-buy, non-trading minutes and bid-ask spread;
 $Extereme_{DAX}$ is equal to 1 for peak observation, -1 for bottoms, and 0 for non-extremes

TABLE 5
EXTREME POINTS OF DAX PRICE IN THE FUTURES LIQUIDITY - BASED
MULTINOMIAL MODEL

Lag	P/B	Volume		SellBuy	NonTrMin	BidAsk	Resid
Panel I: Univariate analysis							
0	P	-0.00002 *** <i><0.0001</i>					
0	B	0.00001 ** <i>0.0517</i>				0.0113 *** <i>0.0148</i>	
1	P	-0.00002 ** <i>0.0407</i>					
1	B	0.000009 ** <i>0.0279</i>		14.56 *** <i>0.0067</i>		-0.00546 *** <i>0.0005</i>	
3	P			-8.70 *** <i>0.0031</i>			
3	B			13.36 *** <i>0.0022</i>			
5	P	-0.00002 * <i>0.0891</i>					
5	B			-7.0919 *** <i>0.0198</i>			
Panel II: Multivariate analysis							
0	P	-0.00006 ** <i><0.0001</i>			-0.783 * <i>0.0572</i>		
0	B					0.0103 * <i>0.0763</i>	
3	P			-50.54 *** <i>0.0122</i>	-0.0341 *** <i>0.0004</i>		
3	B			13.60 *** <i>0.0049</i>			
5	B				-0.00666 * <i>0.0697</i>		
Panel III: Residuals							
0	B						-155.9 ** <i>0.0527</i>

P/B* - peak or bottom

In panel I and panel II, the results of multinomial analysis of the possibility to explain extreme values in the DAX returns trend (peaks and bottoms). Peaks are coded with “1”, bottoms with “-1” and non-extreme observations with “0” for the purpose of the analysis. There are contemporaneous and lagged (1,3, and 5 days) models. The presented coefficients are obtained from the model with Taylor series (linearization) that are generally consistent with simple multinomial model results (not tabulated). The coefficients in bold font are also significant in Jackknife models. We only show statistically significant results. In panel III, DAX return extremes are regressed on residuals from the bivariate model (Table 3, Panel I)

We extend the analysis further to the following multivariable multinomial logit model:

$$Extereme_{DAX} = \beta_0 + \beta_1 Volume_{FDAX} + \beta_2 SellToBuy_{FDAX} + \beta_3 NonTradeMin_{FDAX} + \beta_4 BidAsk_{FDAX} + \epsilon \quad (6)$$

The results of the multinomial logit analysis are presented in Table 5. To conserve space and keep focus on the most important results, we only report estimates of the models that yield statistically significant coefficients. Volume is found to be a significant predictor of both bottoms and peaks of the index: its explanatory power persists in models with explanatory variables lagged up to 5 days, as well as in the multivariable model. Bid-Ask Spread is a strong explanatory variable for contemporaneous index bottoms that maintains its power in the one day lagged model as well.

The methodology for survey data helps overcome the limitations of our fairly small sample. Taylor series (linearization) provides biased, but consistent estimates based on representation of a function obtained from function’s derivatives. The simple form of Jackknife method implemented in our study, removes one observation at a time to form a new subsample. The estimation continues until the end of the sample repeating as many times as many observations are in a sample. Linearization and Jackknife are among the most popular variance estimation techniques for survey data. Rao and Wu (1988), for instance, find that both methods are more stable than bootstrap when dealing with survey samples under severe conditions. Our estimates in the first two panels of Table 5 are a combination of the results obtained using the two above methods. The estimated coefficients in bold font are obtained with Linearization method and are statistically significant under both Linearization and Jackknife. The estimated coefficients in regular font are also obtained with Linearization and are only significant under this method; they are statistically insignificant when Jackknife is applied. Since we use the simplest - “-1 observation” - version of Jackknife, it is very restrictive and estimates are more limited than when using another method. We still trust the estimates obtained with Linearization because they support our early standard multivariate regression results (not tabulated to conserve the space).

The presence of outliers around extreme points in the volume-based regression model (results presented in section 4.1) suggests that volume measures might only explain variation in return on average. However, the unexplained portion of variation may provide a better insight into formation of extreme points of index return. We obtain the residuals $\hat{\epsilon}$ from the regression (2) and run a multinomial logit regression of the peak/bottom identifier on those residuals:

$$Extereme_{DAX} = \beta_0 + \beta_1 \hat{\epsilon} + u \quad (7)$$

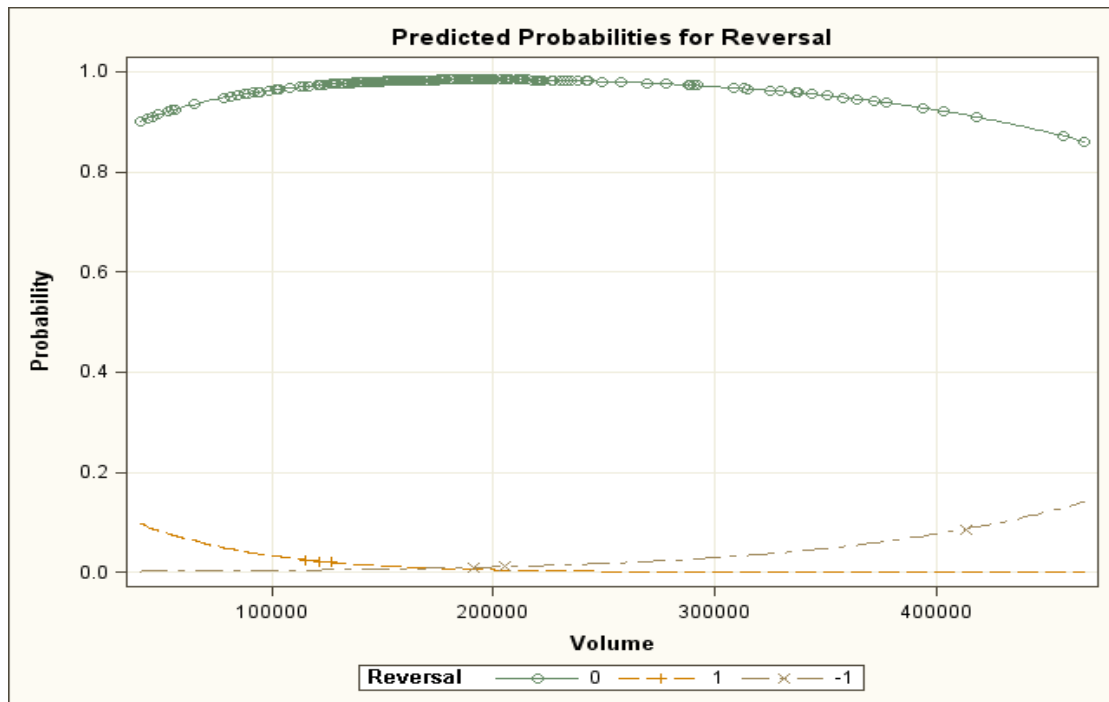
The residuals of the model, based on the futures volume measures, specification (2) are statistically significant explanatory variable for the index return bottoms. Note, the bid-ask spread, which was omitted in specification (2), is also a statistically significant predictor of the index return bottoms.

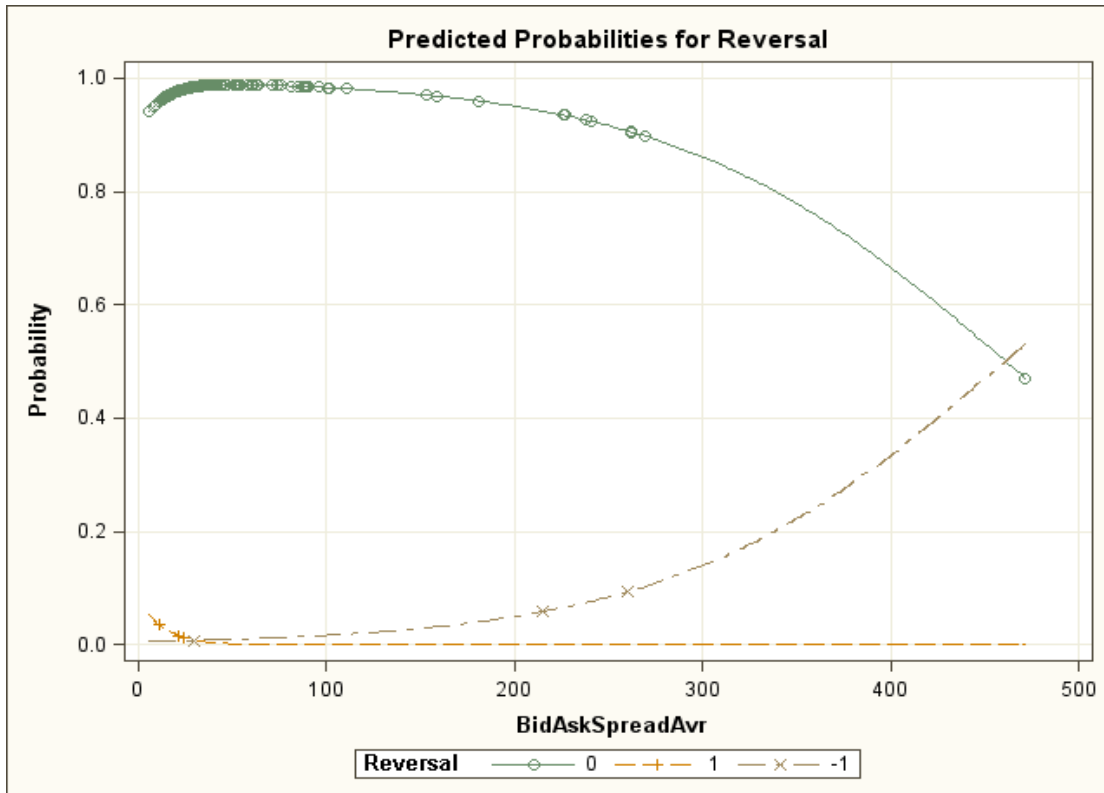
Additional results (not tabulated), including 5-day lagged multivariable model and models based on unsigned extreme variables (peaks and bottoms marked as 1 and other observations - as 0) generally support the above presented findings. These untabulated results are available from the first author.

Via multinomial logit model, we are able to identify statistically significant predictors of the extreme returns. However, we are unable to specify immediately neither direction nor size of the effect (economic significance) as multinomial model coefficients are not directly interpretable. There is a number of ways to

approach the problem of detailed analysis of multinomial logit results. We choose interpretation with plots of predicted probabilities, as suggested by Long (1997). Figure 1 illustrates the effect of two statistically significant explanatory variables – trading volume and bid-ask spread – and shows probability of index peak or bottom depending on the values of predictors. The probability of a peak monotonically declines as futures’ trading volume grows, while a probability of a bottom increases. The empirical peak data points are concentrated in the proximity of just below 200,000 zone, while bottoms reside in approximately 200,000 and higher section. A probability of a peak, in relation to bid-ask spread, gets flat at about 0, after passing the lowest range of bid-ask spread values. This result is supported by the multinomial logit estimates based on a spread that failed to identify any significant relationship in the case of peaks. However, we can clearly see that probability of bottoms increases steeply as bid-ask spread grows wider, reaching around 50 percent towards the end of the available data range.

FIGURE 1
PROBABILITY OF THE EXTREMES IN DAX RETURNS DEFINED BY THE FUTURES’ LIQUIDITY MEASURES





Change Points in the Relationship of Futures' Liquidity and Index Returns

We have learned so far, that multiple dimensions of futures' liquidity are strongly correlated with underlying asset returns, and that relationship is non-monotonic. Thus, slopes of regression lines describing relationship between DAX index returns and liquidity parameters of its futures are non-constant with identifiable change points.

The Markov Chain Monte Carlo (MCMC) method for the change point analysis is a part of the Bayesian statistics framework, a modern alternative to the traditional frequentist methods (Gamerman and Lopez, 2006). Bayesian models do not rely on asymptotic approximations. Instead, they provide inferences obtained from simulated samples that solve a problem of a wide range of models, including models with missing data and models based on small samples. At the same time, Bayesian analysis relies heavily on prior distributions as a basis for posterior distributions in hypothesis testing. The problem of a prior distribution selection remains one of the central discussion topics in the Bayesian statistics literature. There is no generally accepted selection rule and no ultimately correct prior distribution, but results are significantly affected by the choice of a prior. The prior selection problem must be solved anew for each model. We employ a change point regression model suggested by the leading Bayesian statisticians Carlin et al. (1992) and follow the formulation of Best et al. (1996)

$$y_i \sim \begin{cases} \text{normal}(\alpha + \beta_1(x_i - cp), \sigma^2) & \text{if } x_i < cp \\ \text{normal}(\alpha + \beta_2(x_i - cp), \sigma^2) & \text{if } x_i \geq cp \end{cases}$$

where cp – is a change point, β_1 is a slope before change point and β_2 is a slope after the change point and implement the following diffuse prior distributions:

$$\pi(cp) \sim \text{uniform}(\min(x), \max(x));$$

where $\min(x)$ and $\max(x)$ are observed minimum and maximum values of the independent variable

$$\pi(\sigma^2) \sim \text{uniform}(0,5); \pi(\alpha, \beta_1, \beta_2) \sim \text{normal}(\text{Beta}, \text{SE}(\text{Beta}))$$

where Beta is an empirical estimate of the coefficient and SE (Beta) is a standard error of the coefficient estimate;

We use 5,000 tuning samples and 100,000 MCMC (Markov Chain Monte Carlo) samples in our experiment and scale the variables by approximate standard deviation where

$$\sigma = (X_{max} - X_{min})/4 \quad (8)$$

The approximated standard deviation is about 100,000 for volume and about 100 for spread. The most interesting results of the MCMC change point analysis, based on the above described parameters, are presented in the Figure 2. The change point for the scaled volume was identified at 3.0192 (real volume of 301,920). The slope of the regression line is steeper after the change point and majority of the prediction line and empirical observations lay in the negative returns territory once volume reaches the threshold. Results are consistent with the probability profile for volume in the multinomial logit analysis. On the first graph of Figure 1 we can see that probability of negative extremes increases significantly past volume values of 300,000. The watershed for the scaled bid-ask spread was identified at 1.7469 (real bid-ask spread 174.69). The slope of the line slightly flattens past the change point. However, the regression line is in negative returns zone. Many observations are negative and their absolute values are large; positive observations are much smaller. These results are also consistent with the multinomial logit outcomes (graph 2, figure 2) and demonstrate that when bid-ask spread climbs higher than 174.69, the returns tend to gravitate to the bottom of their observed range.

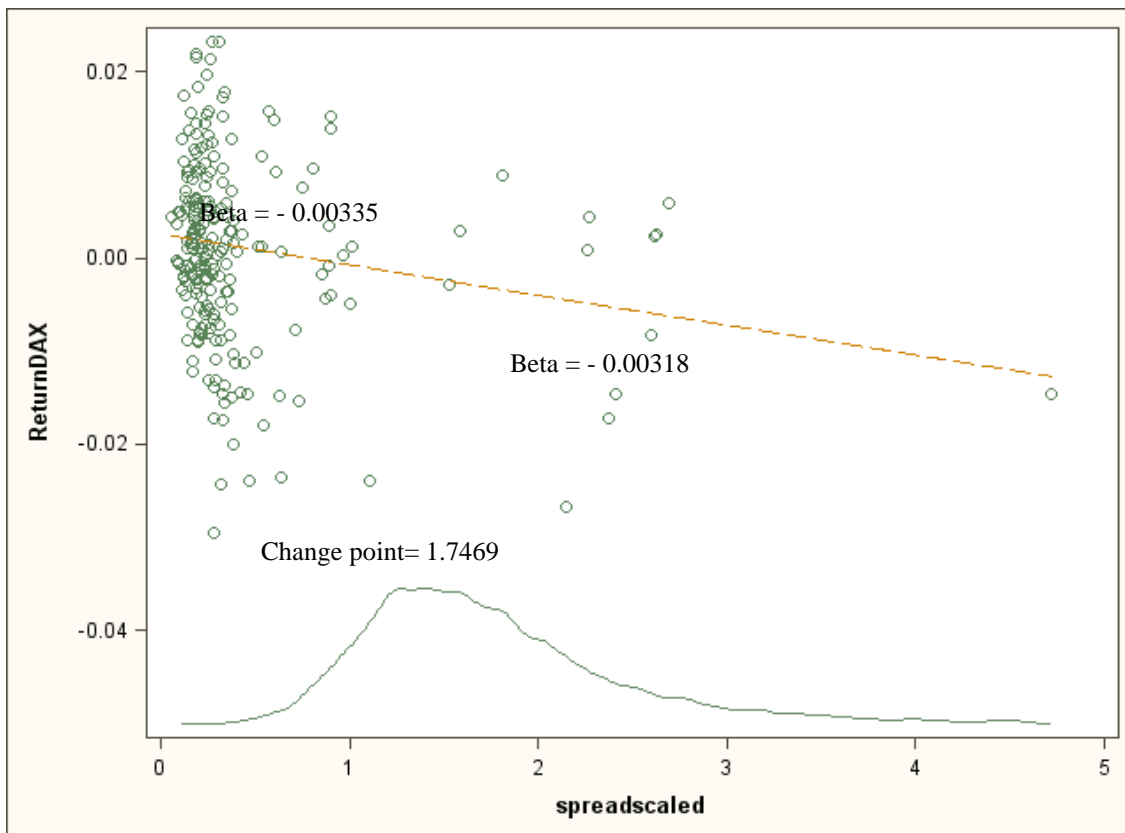
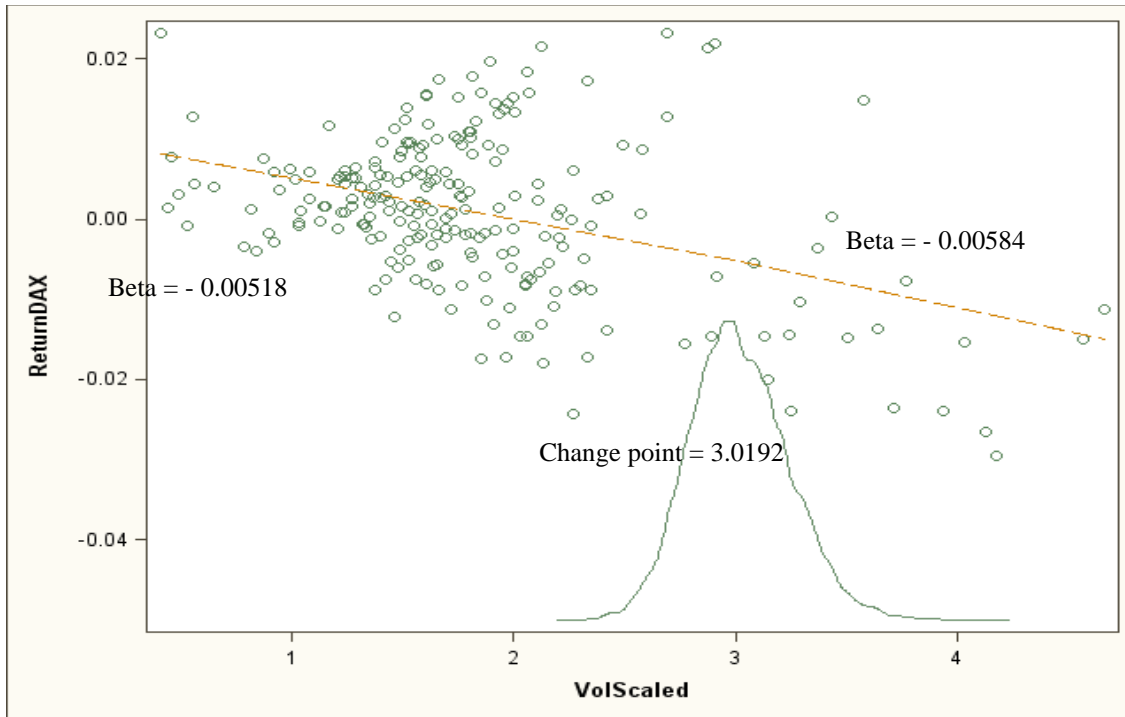
CONCLUSIONS

Our findings add new liquidity-based measures and novel methodology to the asset pricing toolset. The models uncover fresh aspects of price discovery by integrating two important strands in asset pricing literature: studies of a connection between derivative and equity markets, and research on liquidity, priced into assets. The analysis of the correlation between the German equity index DAX and intraday liquidity measures of its futures highlights strong explanatory power of derivative's liquidity measures for the underlying asset returns. Our findings demonstrate that volume-related measures of DAX futures liquidity, bid-ask spread, and number of non-trading minutes in a day, a measure of trading flow consistency, explain a significant portion of DAX returns variability and predict extreme values of the returns.

We show that various liquidity measures capture different aspects in price discovery; those measures complement each other in explaining variability of returns. An array of DAX futures liquidity variables, derived from the high-frequency historical records of Eurex trading, explain up to 20% of variation in DAX returns in our sample. Advanced variable selection methodology assists in building a model with most relevant predictors. The outlier analysis in the trading volume-based regression captures the association of bid-ask spread with extreme values of DAX returns.

Finally, Bayesian change point analysis of trading volume and bid-ask spread of the futures confirms non-monotonic relationship in regressions and identifies the threshold values of bid-ask spread and trading volume that suggest likely positive and likely negative returns. We establish predictability of underlying equity return extremes by intraday liquidity measures in the futures market.

FIGURE 2
CHANGE POINT IN RELATIONSHIP BETWEEN DAX FUTURES LIQUIDITY AND
INDEX RETURNS



ACKNOWLEDGEMENTS

We appreciate valuable feedback received from the conference session participants and discussants at the EFA Annual Meeting and MWDSI Annual Meeting, and from the seminar participants at Kent State University. We are thankful for the comments and suggestions provided by Elyas Elyasiani, Mark Holder, and Lawrence Kryzanowski. All errors and omissions remain our own.

REFERENCES

- Acharya, V., & Pedersen, L.H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, 77(2), 375–410.
- Alessi, L., & Detken, C. (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *European Journal of Political Economy*, 27(3), 520–533.
- Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56.
- Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2), 223–249.
- Amihud, Y., Mendelson, H., & Lauterbach, B. (1997). Market microstructure and securities values: Evidence from the Tel Aviv Stock Exchange. *Journal of Financial Economics*, 45(3), 365–390.
- Austin, P.C., & Tu, J.V. (2004). Bootstrap Methods for Developing Predictive Models. *The American Statistician*, 58(2), 131–137.
- Barnett, V., & Lewis, T. (1978). *Outliers in statistical data*. New York, NY: John Wiley & Sons.
- Belsley, D.A., Kuh, E., & Welsch, R.E. (2005). *Frontmatter In: Regression diagnostics: Identifying influential data and sources of collinearity*. Hoboken, NJ: John Wiley & Sons, Inc.
- Bernanke, B., & Gertler, M. (2000). *Monetary policy and asset price volatility*. No. w7559. National bureau of economic research.
- Best, N.G., Spiegelhalter, D.J., Thomas, A., & Brayne, C.E.G. (1996). Bayesian Analysis of Realistically Complex Models. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(2), 323–342.
- Booth, G.G., So, R.W., & Tse, Y. (1999). Price discovery in the German equity index derivatives markets. *Journal of Futures Markets*, 19(6), 619–643.
- Carlin, B.P., Gelfand, A.E., & Smith, A.F.M. (1992). Hierarchical Bayesian Analysis of Change-point Problems. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2), 389–405.
- Chakravarty, S., Gulen, H., & Mayhew, S. (2004). Informed Trading in Stock and Option Markets. *The Journal of Finance*, 59(3), 1235–1258.
- Chan, K., Chung, Y.P., & Fong, W.M. (2002). The informational role of stock and option volume. *Review of Financial Studies*, 15, 1049–1075.
- Cook, R.D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19, 15–18.
- Cremers, M., & Weinbaum, D. (2010). Deviations from Put-Call Parity and Stock Return Predictability. *Journal of Financial and Quantitative Analysis*, 45(2), 335–367.
- Czuderna, K., Riedel, C., & Wagner, N. (2015). Liquidity and conditional market returns: Evidence from German exchange traded funds. *Economic Modelling*, 51, 454–459.
- Datar, V.T., Naik, N.Y., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2), 203–219.
- Fama, E., & French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2), 427–465.
- Gamerman, D., & Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference* (2nd edition). Chapman & Hall CRC.
- Hasbrouck, J. (2003). Intraday Price Formation in U.S. Equity Index Markets. *The Journal of Finance*, 58(6), 2375–2400.
- Hoerl, A.E., & Kennard, R.W. (1970). Ridge Regression. Applications to non-orthogonal problems. *Technometrics*, 12, 55–67.

- Karpoff, J.M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1), 109–126.
- Knez, P.J., & Ready, M.J. (1997). On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions. *The Journal of Finance*, 52(4), 1355–1382.
- Lim, K.G. (1992). Arbitrage and price behavior of the Nikkei stock index futures. *Journal of Futures Markets*, 12(2), 151–161.
- Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of Financial Economics*, 82, 631–671.
- Long, J.S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications, Inc.
- MacKinlay, A.C., & Ramaswamy, K. (1988). Index-futures arbitrage and the behavior of stock index futures prices. *Review of Financial Studies*, 1(2), 137–158.
- Marshall, B.R., & Young, M. (2003). Liquidity and stock returns in pure order-driven markets: Evidence from the Australian stock market. *International Review of Financial Analysis*, 12, 173–188.
- Martinez, M.A., Nieto, B., Rubio, G., & Tapia, M. (2005). Asset pricing and systematic liquidity risk: An empirical investigation of the Spanish stock market. *International Review of Economics and Finance*, 14, 81–103.
- Merton, R.C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41, 867–887.
- Miller, M.H., Muthuswamy, J., & Whaley, R.E. (1994). Mean Reversion of Standard & Poor's 500 Index Basis Changes: Arbitrage-induced or Statistical Illusion? *The Journal of Finance*, 49(2), 479–513.
- Pan, J., & Poteshman, A.M. (2006). The Information in Option Volume for Future Stock Prices. *Review of Financial Studies*, 19(3), 871–908.
- Pastor, L., & Stambaugh, R.F. (2003). *Liquidity Risk and Expected Stock Returns*. NBER Working Paper Series, 8462.
- Rao, J.N.K., & Wu, C.F.J. (1988). Resampling inference with Complex Survey Data. *Journal of American Statistical Association*, 83(401), 231–241.
- Smith, V.L., Suchanek, G.L., & Williams, A.W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica: Journal of the Econometric Society*, pp. 1119–1151.
- Sorokina, N.Y., Booth, D.E., & Thornton, J.H. (2013). Robust methods in event studies: Empirical evidence and methodological implications. *Journal of Data Sciences*, 11(3), 607–621.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.
- Wang, H., & Leng, C. (2007). Unified LASSO Estimation by Least Squares Approximation. *Journal of the American Statistical Association*, 102(479), 1039–1048.
- Xing, Y., Zhang, X., & Zhao, R. (2010). What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? *Journal of Financial and Quantitative Analysis*, 45(3), 641–662.
- Yadav, P.K., & Pope, P.F. (1990). Stock index futures mispricing: profit opportunities or risk premia? *Journal of Banking & Finance*, 18(5), 921–953.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429.