

What Factors Increase the Risk of Incurring High Market Impact Costs?

Jacob A. Bikker* Laura Spierdijk† Pieter Jelle van der Sluis‡

July 23, 2007

Abstract

This paper applies quantile regression to assess the factors that influence the risk of incurring high trading costs. Using data on the equity trades of the world's second largest pension fund in the first quarter of 2002, we show that trade timing, momentum, volatility, and the type of broker intermediation are the major determinants of the risk of incurring high trading costs. Such risk is increased substantially by either high or low momentum and by strong volatility. Moreover, agency trades are substantially more risky in terms of trading costs than similar principal trades. Finally, we show that the quantile regression model succeeds well in forecasting future trading costs.

Keywords: market impact costs, extreme trading costs, pension funds, quantile regression

JEL classification: G11, G23, C53

*Supervisory Policy Division, Strategy Department, De Nederlandsche Bank, P.O. Box 98, NL-1000AB Amsterdam, The Netherlands. Phone: +31 20 524 2352. Email: J.A.Bikker@dnb.nl.

†Corresponding author. Department of Econometrics, Faculty of Economics, University of Groningen, P.O. Box 800, NL-9700 AV Groningen, The Netherlands. Phone: +31 50 363 5929. Fax: +31 50 363 3720. E-mail: l.spierdijk@rug.nl.

‡Vrije Universiteit Amsterdam, Department of Finance and Financial Sector Management. Also affiliated to Investments Research, ABP Investments, P.O. Box 75753, NL-1118ZX Schiphol, The Netherlands. Phone: +31 20 405 4343. Email: pieterjelle.vander.sluis@Abpinvestments.nl. The views expressed in this paper are not necessarily shared by DNB and ABP Investments or its subsidiaries.

1 Introduction

Since in efficient markets, stock prices move in response to the release of new information, trading itself may cause prices to be revised. Loosely speaking, a buy trade tells the market that a stock is undervalued, and similarly, a sell trade indicates that a stock is overvalued. Market participants observe the information conveyed by trading and adjust their perceptions accordingly, which results in price movements. Other reasons for stock prices to move in response to trading are e.g. demand and supply imbalances and liquidity effects.

When an investor sends an order to the market, it is usually not executed directly. For a large trade done by e.g. an institutional investor, it usually takes some time to find a counter party. Since large trades convey information, the price at trade execution is generally different than at trade initiation. Market impact costs occur when price effects cause execution prices to be less favorable than benchmark prices.

A stock with a high gross return may yield a relatively low net return if trading costs are high. Similarly, a stock with a low gross return can yield a relatively high net return given low transactions costs. Stated differently, trading costs affect the net return on an investment and thus influence optimal portfolio holdings. Therefore, trading costs play an important role in portfolio management.

The starting point of this paper is our observation that a relatively small group of equity trades causes the major part of market impact costs. For the equity trades investigated in this paper, executed by the world's second largest pension fund, we find that only about 10% of the trades causes approximately 75% of total market impact costs. Consequently, reducing or avoiding the trading costs of even a few expensive trades would result in substantial savings on total costs. Therefore, it is important for the investor to know what trade properties and market circumstances increase the risk of incurring extreme trading costs.

The existing literature on trading costs usually applies a regression approach to analyze the average impact of certain determinants on market impact costs. Var-

ious trade-, exchange- and stock-specific characteristics have been shown to affect market impact costs. See Keim and Madhavan (1998) and Bikker et al. (2007) for an overview. However, the usual regression-based method is not suitable to assess the factors that cause trading costs to be extreme. That is, the traditional setup focuses on the average impact of certain trade properties and market circumstances on trading costs rather than on the extreme effects that these determinants occasionally have. Analyzing the variance of market impact costs (see Bikker et al. (2007)) suffers from the disadvantage that variance is a one-sided risk measure that does not distinguish between extremely low and extremely high trading costs. Therefore, this paper takes a different view, focusing on the determinants of extremely high trading costs and using quantile regression to assess these factors. With the flexible quantile regression approach, we can explicitly determine the impact of trade characteristics and market conditions on the $100\tau\%$ most expensive trades, where τ can take any value in the interval $(0, 1)$. Besides its flexibility, another advantage of quantile regression is its relative robustness to outliers in the dependent variable, which are abundantly present in the data used in this paper.

In practice, a major motive for estimating market impact models is their use for prediction purposes in trading cost management. Therefore, this paper also addresses the issue of forecasting future market impact costs. Forecasts of transaction costs, particularly market impact costs, can be integrated in the portfolio construction at the same stage of the optimization process where risk and return forecasts are applied. In this way the components return, risk, and costs are optimized simultaneously. We evaluate the performance of the quantile regression model in terms of its forecasting quality.

The data used in this paper cover the global equity trades executed by the world's second largest pension fund (the Dutch 'Algemeen Burgerlijk Pensioenfonds' (ABP)) during the first quarter of 2002. This unique data set includes a vast collection of trade, exchange, and stock specific characteristics corresponding to each trade.

Previous research has established important roles for trade style and variables related to trade difficulty (such as market capitalization and trade size) in explaining market impact costs. We find that various trade characteristics and market conditions increase the risk of incurring high trading costs. In particular, conditions of high and low momentum and strong volatility make extreme market impact costs more likely. The type of broker intermediation also appears to be an important risk factor. Agency trades are more risky in terms of trading costs than similar principal trades, while trade timing also increases risk substantially. Moreover, we show that accurate prediction of market impact costs poses a difficult challenge. Therefore, we take a different view and focus on predicting the entire distribution of market impact costs instead of expected values. We show that quantile regression – not applied before in this field – is able to successfully forecast this distribution and outperforms the traditional regression approach in terms of predictive power.

The setup of this paper is as follows. Section 2 introduces the concept of market impact costs and presents the data used in this paper. This leads us to Section 3, which explains the quantile regression approach and subsequently discusses the estimation results for the quantile regression model. Moving forward to Section 4, we discuss how the quantile regression model can be used to form expectations about future market impact costs on a stock. Subsequently, we assess the forecast quality of the quantile regression model. Finally, Section 5 summarizes and concludes.

2 Preliminary data analysis

The data used in this paper comprise all internally managed equity transactions of the world’s second largest pension fund, ABP (‘Algemeen Burgerlijk Pensioenfonds’), during the first quarter of 2002. ABP has some 2.4 million clients and an invested capital of approximately 200 billion euro¹, corresponding to one third of total Dutch pension fund assets. In the period under consideration there were ten internal funds

¹This is the total invested capital in July 2007.

in ABP's equity group, apart from the externally managed funds. The unique data set contains detailed information on 3,721 equity trades executed during the first quarter of 2002, with a total transaction value of 5.7 billion euro. The ten internally managed equity portfolios in our sample had a total value of 20 billion euro. Of all trades, 1,962 were buys and 1,759 were sells executed in Europe, the United States, Canada, and Japan.

Trading process

Before turning to the trading process at ABP, we distinguish three types of trade. A principal trade is a transaction between the pension fund and the broker, in which case the broker buys or sells stocks from or to the pension fund at a predetermined price. Hence, the risk is transferred to the broker. The broker takes on the other side of the trade and tries to execute the trade in the open market. An agency trade is a trade between the pension fund and a counterparty, where the broker acts solely as an intermediate party. Thus, an agency trade involves two clients of the brokerage firm, one of which is the pension fund. The term 'single' trades is applied to difficult trades that are done separately, either with or without packages of other stocks. In the case of single and agency trades the risk resides with ABP. The broker represents the client (ABP) and acts in the client's best interest.

For all trades the trading process during the first quarter of 2002 was as follows. A portfolio manager formed his or her portfolio. Subsequently, he or she approached a trader at ABP. Together they discussed the proposed trade. In most cases the trader would single out some part of the trades (say 10%) for reasons of perceived cost reduction and would execute these elsewhere in the market as agency or single trades. Next, the trader approached at most two of the large brokerage firms for the remaining trades and revealed some of the characteristics of the trade (volume, US or Europe, quantitative or fundamental, sector decomposition and a judgment on the complexity of the trade). The choice of brokerage firms was based on the

trader' experience. Only the largest brokers could make competitive offers in the case of principal trades, although sometimes a smaller one had an edge in particular market segments, for example Japan. Based on the characteristics of the trade, the broker made an offer for a principal trade. The the broker's offer was compared to others and to the trader's own systems and experience. If the offer was acceptable, then a principal trade was executed. Otherwise, the trade was executed as an agency trade.²

Data and definition

For each transaction the data provide the execution price. Also provided is the price of the stock just before the trade was passed on to the broker, which is the price at which the portfolio managers allow the traders to trade. Moreover, the data also show the time the trade was submitted to the broker and the time it was executed. Trades that were split up into several subtrades are considered as a single trade if the decision to split up the trade was taken by a trader at ABP. The data contain about 0.5% of such 'trade packages'. Orders split up by portfolio managers are treated as individual trades, since it is not known whether the traders eventually split up the trade the same way the portfolio managers did. Additionally, the data include detailed information on several trade, exchange, and stock specific characteristics. Table 1 provides a complete list of the variables in the data set, including their abbreviations and definitions. For a complete description of the data, we refer to Bikker et al. (2007).³

²For a detailed exposition of the trading process at ABP, we refer to Bikker et al. (2007).

³The data set was created on the basis of the post-trade analysis provided by ABP, in combination with additional data from Factset and Reuters. The information on the characteristics of the exchanges under involved were obtained from the World Federation of Exchanges and the exchanges themselves.

Measuring market impact costs

In order to measure market impact costs of ABP trades, a benchmark price has to be chosen. The literature provides many suggestions for this choice. A commonly used same-day benchmark is the volume-weighted average price calculated over all transactions in the stock on the trade day. Pre-execution benchmarks are based on the opening price on the same day or the closing price on the previous day. Finally, post-execution benchmarks take the closing price of the trading day or the opening price on the next day as reference price, ensuring that the temporary price impact has disappeared from the benchmark. For a discussion of these approaches, see Collins and Fabozzi (1991) and Chan and Lakonishok (1995). This paper opts for the pre-execution benchmark, in line with e.g. Wagner and Edwards (1993). More precisely, we take the price at the moment the order was passed to the broker as benchmark and correct for market-wide price movements during the trade, as do Chan and Lakonishok (1995, 1997). The MSCI World Industry Group Indices are used as a proxy for these market movements. Thus, for a buy transaction in stock i at time t market impact costs (C_{it}^B) are measured as

$$C_{it}^B = \underbrace{\log(P_{it}^{exe}/P_{it}^{pt})}_{\text{price impact}} - \underbrace{\log(M_{it}^{exe}/M_{it}^{pt})}_{\text{market wide price movements}}, \quad (1)$$

where P_{it}^{exe} and P_{it}^{pt} denote the execution and pre-trade price of stock i at time t , respectively. M_{it}^{exe} and M_{it}^{pt} denote the value of the MSCI industry group index corresponding to stock i at the time of the execution of the trade and at the pre-trade time, respectively. Similarly, the market impact of sells is defined as

$$C_{it}^S = \log(P_{it}^{pt}/P_{it}^{exe}) - \log(M_{it}^{pt}/M_{it}^{exe}). \quad (2)$$

For both buys and sells, positive market impact implies that a trade has been executed against a price worse than at the moment of trade initiation.

Some sample statistics

Table 2 reports sample means, standard deviations, medians and quantiles of market impact costs for buys and sells. The sample statistics are provided both on a principal-weighted basis and unweighted. We obtain the principal-weighted statistics by weighting each observation by the euro value of the trade, so that larger trades contribute more to, for instance, the average market impact than smaller ones. Average market impact costs of buys equal 19.6 basis points (bp) and those of sells 29.7 bp. The large quantiles in the right tail of the distribution are the trades that really matter in terms of trading costs and that play an important part in cost management. Their contribution to the spread of market impact costs around the mean or median is substantial. These expensive trades take up a central position in this paper. Table 2 shows that, there are relatively many trades that actually result in profits instead of losses. This is emphasized by further calculations that show that only about 10% of the trades causes approximately 75% of total market impact costs. The unweighted average market impact costs for buys are of the same magnitude as the weighted ones, but for sells we observe a much lower value (even slightly negative). Additionally, Table 2 shows that the unweighted quantiles are often of considerably smaller magnitude than the weighted ones, which implies that a few large trades with high market impact costs determine the major part of the weighted average. The asymmetry in market impact costs of buys and sells is also found in other studies, see e.g. Kraus and Stoll (1972), Holthausen et al. (1987, 1990), Chan and Lakonishok (1993, 1995), and Keim and Madhavan (1996, 1997).

To give an idea of the transactions executed by ABP, we report some sample statistics. The average trade size for buys (sells) is more than 70,000 (84,000) shares and the average value of a trade equals almost 1.5 (1.6) million euro. Expressed as a percentage of daily trading volume and shares outstanding, the average trade size of buys equals 4.29% and 0.02%, respectively. For sells these percentages are 3.41% and 0.02%. Commission averages about 12 bp for both buy and sell transactions.

The average buy (sell) takes about 4 (4.5) hours to be completed.

3 Model selection, estimation, and interpretation

This section briefly introduces the notion of quantile regression and subsequently discusses model selection and estimation. Finally, it turns to the interpretation of the estimation results.

3.1 Quantile regression

Koenker and Bassett (1978) introduced the method of quantile regression in the literature. Many applications followed, resulting in a vast literature on the subject. Buchinsky (1998) provides an excellent survey. Furthermore, Koenker and Hallock (2001) give many practical suggestions. In this subsection we briefly review the main properties of quantile regression and explain how it differs from the classical regression model. To clarify the exposition, this subsection focuses on market impact costs as the dependent variable and trade characteristics as the covariates. However, they can be replaced by any other variables.

The focus of the classical regression model is on the determinants of *expected* trading costs. Quantile regression, however, has a broader scope and can be used to assess the factors that influence the $100\tau\%$ most expensive trades, where τ can take any value in the $0-1$ range. Consequently, quantile regression does not yield a single estimate, but a set of estimates covering the entire range from cheap to expensive trades. Just as the linear regression model assumes that conditional expected trading costs C , given a K dimensional vector of covariates X , equal $X\alpha$ (with α a K -dimensional vector of coefficients), the quantile regression model assumes that the τ th conditional quantile of C , given the covariates X , equals $X\beta_\tau$ (with β_τ a K -dimensional vector of coefficients).

Not only does the quantile regression approach allow direct analysis of the entire distribution of trading costs, it is also considerably more flexible than the classical

regression model in capturing the way market conditions and trade characteristics affect the cost distribution. The usual regression approach implicitly assumes that any determinant of trading costs affects these costs only through the mean and – in a heteroskedastic regression model – through the variance. By contrast, the quantile regression approach allows the impact of trade characteristics and market conditions on trading costs to depend on the trading cost level. For a more technical discussion of quantile regression, see Appendix A.

3.2 Model selection

Using Roger Koenker’s Quantreg package (version 3.90) for R, we estimate the quantile regression process over the range from cheap to expensive trades using the Barrodale and Roberts algorithm (see Koenker and d’Orey (1987, 1994)). We obtain standard errors from Powell (1986)’s kernel estimator for the covariance matrix. In line with the literature, we estimate separate models for buy and sell trades.

Initially, we estimate a full quantile regression model for both buys and sells, containing all explanatory variables described in Section 2 and summarized in Table 1. The corresponding estimation results suggest that the coefficients of several (mostly dummy) variables are not significant at any range of quantiles and can be left out. To test this in a formal way, we use a Wald test to conduct a model selection procedure from general to specific (see Koenker and Portnoy (1999)). As expected, the Wald test shows that the coefficients of several variables are jointly insignificant. We delete these variables from the initial model to obtain a more parsimonious specification.

Figures 1 and 2 display the coefficients of the variables included in the final model specification for buys over the entire range of quantiles. Similarly, Figures 3 and 4 show the coefficients for sell trades. These figures do not only display the values of the estimated coefficient over the entire range of quantiles (solid curve), but also depict 95% confidence bounds (light-shaded area). Additionally, ordinary least squares coefficients obtained from the standard regression model (solid black

line) and corresponding 95% confidence intervals (dashed black lines) obtained from the classical linear regression model are plotted to facilitate visual comparison of quantile regression with the traditional regression approach.⁴

3.3 Model interpretation

This subsection provides some economic explanations for the results obtained with the quantile regressions as displayed in Figures 1-4.

First, we explain briefly how to interpret the plots in Figures 1-4. Each individual plot depicts the impact of a variable on the range of quantiles of market impact costs, starting with the lowest quantile on the left and ending at the highest quantile on the right. Throughout, we will zoom in on the factors that affect the right tail, since we are most interested in the trade characteristics and market conditions that affect the *risk* of incurring extreme trading costs.

Momentum

Momentum has a significantly positive effect on almost all quantiles of the costs distribution of buys and a significantly negative influence on, particularly, the right tail of the cost distribution of sells. When momentum increases, the liquidity costs of a buy transaction will become larger in order to convince more stock owners to sell their shares. Similarly, when momentum drops, liquidity costs will decrease to make it more attractive for market participants to buy stocks. Also, a buying (selling) trend in the market may reveal the presence of good (bad) news, so an increase (decrease) in momentum will lead to an increase in the information content of a buy (sell). Together, these two effects explain the positive (negative) impact of momentum on the market impact costs of buys (sells).

The U-shaped form of the quantile plot for buys indicates that, in periods of high

⁴As a robustness check, we have also calculated confidence bounds for the quantile regression estimates using Chernozhukov (2005)'s extremal quantile regression approach (for which we used the R code available from the author's web site). The resulting confidence bounds were very close to the current ones. Therefore, we do not depict them in Figures 1-4.

momentum, extremely high trading costs for buy trades are more likely, whereas exceptionally low costs are less probable. For sells, momentum also has substantial influence on the upper tail of the cost distribution. The inversely U-shaped form of the quantile plot implies that the risk of incurring extremely high trading costs on sells is higher in periods with a selling trend.

Our regression results show that when momentum prior to trading (as measured by the 5-day volume weighted return) increases, trading costs increase as well. This means that higher returns are to some extent offset by higher market impact costs. The setting of our study is clearly different from Korajczyk and Sadka (2004) and Lesmond et al. (2004), who focus on returns net of trading costs. Nevertheless, the positive relation between momentum and trading costs that we establish, could possibly explain why these two papers find that positive returns of momentum trading strategies disappear when trading costs are taken into account.

Volatility

For both buys and sells, the narrow confidence bound around the coefficients of volatility points to a high level of significance. Price volatility only has a scale effect on trading costs. The cost distribution is more dispersed in periods of high volatility; i.e. extremely low and high market impact costs are more likely under such circumstances. We explain this from the observation that more volatile stocks experience stronger idiosyncratic price fluctuations, causing greater dispersion in market impact costs.

Market capitalization

The literature usually establishes a negative relation between market capitalization and market impact costs: the more liquid a stock, the lower the costs of trading. See e.g. Keim and Madhavan (1997).

The results established for sells are generally in line with the literature. The risk of incurring extremely high trading costs with selling is lower for large cap

stocks than for small cap ones. By contrast, the lower tail is not significantly affected by market capitalization. The picture for buys is completely different. Market capitalization only affects the quantiles in the lower half of the cost distribution: the probability of incurring extremely low values of market impact costs on buys decreases with market capitalization.

Trade size

The quantile regressions show that the impact of relative trade size on market impact costs is virtually always positive and fairly constant over the range from cheap to expensive trades.⁵ The coefficients of the quantile regression estimates generally fall within the confidence intervals corresponding to the classical regression model. The only exception relates to sells in the upper tail. Hence, trade size mainly exerts a location effect on trading costs. The positive effect of trade size on trading costs is in line with the literature: large trades generally put more pressure on liquidity and convey more information than smaller ones, resulting in higher price impact. See e.g. Easley and O'Hara (1987).

Agency/single and principal trades

The literature has pointed out that principal trades benefit from the broker's objective to maintain his reputation. The broker will attempt to limit market impact costs on principal trades, resulting in lower trading costs for principal trades than for agency/single trades. See Smith et al. (2001).

The important role of broker intermediation is underlined by the narrow confidence bound around the respective coefficients in the models for both buys and sells. For both categories, the type of broker intermediation strongly affects the tails of the cost distribution. For agency/single buys, the coefficient corresponding to the broker variable is significantly positive everywhere except at the lowest quantiles where it is not significant. Hence, the risk of incurring extremely high trading cost

⁵We notice that interaction terms such as the product of relative trade size and volatility are not included in the regression model as their impact on market impact costs is not significant.

levels is higher for agency buys than for similar principal buy trades. For sells, the coefficients of the broker variable are significantly negative at the quantiles in the left tail of the cost distribution and significantly positive in the right tail. This outcome suggests that compared to principal sell trades, agency/single sells carry both a high risk of incurring very high trading costs and a high probability of encountering very low market impact costs.⁶

Quantitative and fundamental funds

Chan and Lakonishok (1993, 1995) and Keim and Madhavan (1997) find that quantitative funds trade with more urgency than fundamental funds and are willing to pay the price for higher immediacy, resulting in higher market impact costs.

For buy trades the coefficients of the investment style variable are significantly negative everywhere except in the right tail where they are not significant. This means that the probability of incurring very low trading costs is higher than for similar buys executed by fundamental funds. By contrast, the coefficients for sells are significantly positive throughout, with the exception of the left tail where they are not significant. Hence, sells executed by quantitative funds are, in line with theory, more likely to incur extremely high trading costs than comparable sells executed by fundamental funds.

Trade timing and trading venue

The timing of trades has substantial impact on trading costs. For instance, buy trades initiated before or at the opening of the market are more likely to incur exceptionally high trading costs. The finding that the time of the day affects the

⁶We note that the pension fund itself decides whether it wants to trade on an agency/single or principal basis. Clearly, the pension fund's choice for either an agency/single or principal trade may be affected by the expected market impact costs of the trade, which, in turn, is one of the determinants of the initial choice for a specific trade type. This may cause a selectivity or selection bias, see Heckman (1976, 1979). For a detailed survey of the selectivity bias literature, we refer to Vella (1998). When the selection effect is ignored, the resulting estimators may be inconsistent. To assess the possible selectivity effects regarding the choice of trade type, we conducted a similar analysis as Madhavan and Cheng (1997). Using a two-stage estimation procedure we estimated a probit-model to explain the choice for an agency/single or principal trade and a regression model with a correction factor for selectivity effects depending on the probit-specification. We did not establish significant evidence for a selection bias.

costs of trading is in line with Foster and Viswanathan (1993) and McNish and Wood (1992), who establish strong intraday patterns in adverse selection costs and bid-ask spreads, respectively.

The trading venue also plays an important role in explaining trading costs. For example, sells traded on the NYSE are less likely to incur very low values of market impact costs than trades executed elsewhere.

As with the sample statistics in Section 2, we again establish substantial differences between buy and sell trades. Also, the results show that average effects of market conditions and trade characteristics are usually in line with what we would expect from the literature. However, quantile regression depicts the relation between market impact costs and those variables in more detail. In particular, it highlights the influence of the various factors on the risk of incurring extremely high trading costs. In most cases, the impact found in the tails of the cost distribution is significantly different from the average effect.

3.4 Model robustness and goodness-of-fit

Figures 1-4 suggest that the relation between market impact costs and their determinants strongly varies across the range from cheap to expensive trades. Most quantile regression estimates lie at some point outside the confidence intervals of the classical regression model. This shows that the classical regression model is not flexible enough to capture the distribution of the trading costs over the entire range from cheap to expensive trades.

To underpin this conclusion, we test the quantile regression process against the classical and heteroskedastic regression models (see Appendix A) using the Khmaladze (1981) test⁷ proposed by Koenker and Xiao (2002). The null hypothesis of the classical regression model (or the heteroskedastic regression model) is rejected when the joint test statistic exceeds the joint critical level. The test results show that

⁷This test is implemented in the `Quantreg` package in R.

both models are indeed rejected at every reasonable significance level in favor of the quantile regression model.⁸

To get some idea of the goodness-of-fit of the quantile regression process, we calculate the ‘pseudo R^2 ’ corresponding to the quantile regression models for buys and sells. The pseudo R^2 takes values in the interval $[0, 1]$ and is the natural extension of the R^2 found in the classical regression model to the quantile regression approach - we refer to Koenker and Machado (1999) for a more precise definition of this goodness-of-fit measure. Interestingly, the pseudo R^2 -’s displayed as a function of τ (not displayed here to save space) are shaped like a parabola. For buys, the minimum of about 0.10 is attained at the 0.4th quantile. For sells, the lowest value equals 0.14 and is reached around the 0.5th quantile. For both buys and sells the pseudo R^2 peaks at approximately 0.35 when τ reaches the limits of the interval $[0, 1]$. Hence, the best fit is obtained for the most extreme quantiles. In the usual regression model (with the same explanatory variables as considered for the quantile regression process), the R^2 equals 0.18 for buys and 0.24 for sells.

3.5 Relative importance of variables

Figures 1-4 demonstrate that many variables significantly affect the distribution of market impact costs. A question as yet unanswered is which of these variables succeeds best in explaining market impact costs. A complicating factor in answering this question is the fact that not all variables will influence each part of the market impact cost distribution to the same extent. For example, some variables will only affect part of the distribution, for instance the lower tail. Given the scope of this paper, we are most interested in assessing the covariates that are responsible for the extremes of market impact costs, i.e. the variables that substantially influence the right tail of the cost distribution.

The literature on the classical regression model has proposed many ways to assess

⁸The (technical) results of the Khmaladze (1981) test are available from the authors upon request.

the relative importance of explanatory variables, such as the well-known partial R^2 (reflecting the proportion of unexplained variation of the dependent variable that becomes explained with the addition of a covariate to the model). With quantile regression the situation is more complex, as we consider a range of models over the quantiles in the interval $(0, 1)$. In this case it is natural to measure the relative importance of a covariate over a range of quantiles of the dependent variable. Therefore, we generalize the *squared partial correlation* (SPC) as applied in the classical regression model to the quantile regression framework. In the linear regression model, the SPC reflects how much of the variance in the dependent variable that is not associated with any other predictors, is associated with the variance in a particular covariate. The squared partial correlation is calculated as

$$SPC = (R^2 - R_{-i}^2)/(1 - R_{-i}^2), \quad (3)$$

where R^2 is the adjusted R^2 of the full model (containing all explanatory variables) and R_{-i}^2 the adjusted R^2 corresponding to the model without covariate i . Because of the convenient definition in terms of the R^2 , we can easily adjust this measure for the quantile regression model by defining it in terms of the pseudo R^2 (denoted by \tilde{R}^2). As a consequence, the SPC then becomes a function of the quantile $\tau \in (0, 1)$. That is,

$$SPC(\tau) = (\tilde{R}(\tau)^2 - \tilde{R}(\tau)_{-i}^2)/(1 - \tilde{R}(\tau)_{-i}^2). \quad (4)$$

For each explanatory variable in the quantile regression model, we calculate the SPC-'s for a range of quantiles. To obtain an impression of the contribution of each variable to the left and right tail of the cost distribution, we calculate the average SPC over the quantiles $\tau = 0.01, \dots, 0.1$ (left tail) and $\tau = 0.9, 0.91, \dots, 0.99$ (right tail). To assess the impact of each covariate on the center of the cost distribution we calculate the partial autocorrelation at the median ($\tau = 0.5$). Table 3 reports the SPC-'s for each explanatory variable under consideration in, respectively, the

model for buys and sells. The SPC-'s show that a number of variables related to the timing of trades strongly influence the entire cost distribution. Momentum and volatility hardly affect the center of the cost distribution, but substantially influence its tails. The type of broker intermediation also plays an important role in explaining market impact costs, in particular its right tail. Furthermore, the SPC-'s show that the economic importance of some other variables – although statistically significant – is limited.

4 Forecasting future market impact costs

Our model for trading costs can be used to make better informed decisions in portfolio management. As explained by Grinold and Kahn (1999), forecasts of transaction costs, particularly market impact costs, can be integrated in the portfolio construction at the same stage of the optimization process where risk and return forecasts are used (see e.g. Schittenkopf et al. (2002) and Moreno et al. (2005)). In this way the components return, risk, and costs are optimized simultaneously. We note that some of the determinants of market impact costs, such as volatility and momentum, are exogenous to the pension fund. Other determinants, such as the day of the week and the agency-principal dummy, are under the control of the fund and can therefore be integrated in the portfolio optimization process. Alternatively, the market impact cost model can be used to identify those factors (e.g. trade size as a percentage of average daily volume) which contribute most to market impact costs. This knowledge makes it possible to control the active bets in a similar way other investment constraints do so. Also, the cost forecasts can be used to identify potentially expensive trades in terms of market impact costs. Such forecasts have a signalling function in the trade monitoring phase. A different trading strategy can be adopted for trades that are likely to turn out expensive.

As a first step in the use of cost forecasts in the portfolio optimization process, this section focuses on forecasting expected values of market impact costs. We calcu-

late various quantitative forecasting performance measures to determine the extent to which we are able to predict trading costs. Furthermore, we discuss an alternative way to form expectations about future market impact costs and we suggest ways to incorporate forecasts of future trading costs in the trading process.

We compare the model’s forecasting ability not only in-sample, but also for an out-of-sample period. First, we divide the data sample into an in-sample part (the first two months of trades, about 75% of the total sample, say trades $t = 1, \dots, n$) and an out-of-sample part (the final month of trades corresponding to 25% of the sample, say trades $t = n + 1, \dots, n + m$). Next, we estimate the model using only the in-sample data. The in-sample forecasts correspond to the predicted trading costs for the in-sample trades. To obtain forecasts for the out-of-sample trades, we estimate the model using all trades up to rebalancing k . Next, we calculate forecasts for all trades at rebalancing $k + 1$. We repeat this step-wise for each rebalancing k , resulting in m out-of-sample forecasts.

4.1 Forecasting numeric values of market impact costs

The simplest way of forecasting market impact costs is to take the conditional expected value as a forecast of future trading costs. We estimate α in the linear regression model $C = X\alpha + \varepsilon$ by means of ordinary least squares and take $X\hat{\alpha}$ as a forecast of market impact costs. We obtain the final set of regressors X by performing a similar model selection procedure as before. However, we remove the variable trade duration from the set of regressors since trade duration is not known up front. Also, we use only the in-sample data period to perform the model selection. This procedure selects the same variables as before. Figures 5 (a)-(c) display realized and forecasted market impact costs for buys and sells during the out-of-sample period. The difference between realized and forecasted trading costs is largest where these costs take extreme values, while the forecasts seem more accurate when market impact costs are moderate. However, at this point we emphasize that the extreme

cases of trading costs are most relevant from the perspective of cost management. Although the scale of the forecasts (light-colored line) is different from the realized trading costs (dark-colored line), the direction of the forecasts seems to follow realized trading costs quite well. However, the quality of the predictions is difficult to judge visually. Therefore, Table 4 displays various error measures and other quantities that relate to the quality of the forecasts. *Theil's inequality coefficient* and the mean absolute relative error are scale-independent measures for the forecast error and should be as close to zero as possible. The mean squared error is another measure for the prediction error, but it depends on the scale of the data. We also report its decomposition into bias, variance, and covariance percentages which sum up to 100. The bias percentage tells us how far the mean of the forecast is from the mean of the actual series, whereas the variance percentage measures the variation of the forecast relative to the variation of the actual costs. The covariance percentage measures the remaining unsystematic forecasting errors. Ideally, the bias and variance proportions should be small so that most of the discrepancy between forecasted and realized market impact costs is idiosyncratic. The hit ratio counts the percentage of forecasts with the correct sign. Table 4 also displays the 'naive' hit ratio (obtained by assigning each trade to the most likely category). Only for sells in the out-of-sample period is the hit ratio of the quantile model lower than the naive hit ratio. Finally, Table 4 reports the correlation between the forecasted and realized trading costs, which should ideally be as close to one/1 as possible. For more information on the error measures and their definitions, we refer to Appendix B.

The results show that, as expected, the performance of the forecasts is generally better in-sample than out-of-sample. For the in-sample forecasts, the bias proportion is low. Out-of-sample the bias is much higher. Moreover, the in-sample and out-of-sample variance proportions are quite high, although lower than the covariance parts. The correlations between forecasted and realized trading costs reflect the extent to which the model forecasts higher trading costs for stocks that actually do incur high

costs. Both in-sample and out-of-sample, the correlation is significantly positive, reflecting a modest positive relation between forecasted and realized trading costs.

The results demonstrate that it is difficult to obtain accurate forecasts of market impact costs, in particular for trades with extremely low or high trading costs. Since market impact costs reflect the price movements of a stock during trade execution, this difficulty does not come as a complete surprise. Moreover, the out-of-sample period differs substantially from the in-sample months, which also complicates forecasting.⁹

4.2 Forecasting the distribution of market impact costs

This subsection presents an alternative approach and focuses on forecasting the distribution of market impact costs, rather than its expected future value. Since any distribution is completely determined by its quantiles, we forecast the conditional distribution of market impact costs via its quantiles obtained from the quantile regression process. Note that this approach reduces to a Value-at-Risk (VaR) analysis applied to market impact costs. A forecast of – particularly the upper tail of – the conditional distribution of future market impact costs can play a useful role in managing trading costs. For instance, when the 95% conditional quantile of a trade exceeds a certain critical level, the investor can classify the trade as ‘risky’ and decide to manage the trade in a different way in order to avoid high trading costs. Alternatively, a fully integrated approach applies a VaR analysis to returns net of trading costs. Finally, forecasts of conditional quantiles can also be used to construct prediction intervals. For example, a 90% prediction interval is the range of values in which future market impact costs will fall with 90% probability. Instead of forecasting a single value, we can use prediction intervals to forecast a range of values in which trading costs are likely to fall.

Table 5 displays in-sample and out-of-sample estimates of several quantiles (see

⁹In the year 2002, January was bearish and February was quite flat. However, the out-of-sample month of March was bullish.

the columns captioned ‘ $\hat{\tau}$ ’), based on the quantile regressions. Moreover, Figures 5 (b)-(d) depict 90% prediction intervals for realized trading costs based on the 5% and 95% conditional quantiles obtained from the quantile regression model. The prediction intervals are asymmetric and quite broad, which is no surprise given the large dispersion in market impact costs. Furthermore, for 95.2% respectively 91.5% of the out-of-sample observations the trading costs of buys and sells fall within the 90% prediction interval. Thus at first sight, the coverage of the prediction interval during the out-of-sample period is quite accurate, but we need a formal statistical test to confirm this.

The quality of conditional quantile forecasts can be formally evaluated in any of several ways. We opt for an approach that is often used in Value-at-Risk (VaR) analysis. Therefore, at this point, we emphasize the relation between forecasting market impact costs and forecasting VaR. Instead of predicting extreme price depreciations such as in VaR-analysis, we wish to predict extreme cases of market impact costs. Because of this close connection, techniques developed to evaluate the quality of VaR forecasts can also be used in the current setting. Thus, we assess the quality of the forecasted quantiles by calculating the fraction of ‘exceptions’ (the number of times realized trading costs are less than the estimated τ th quantile) to see how close it is to τ , which yields a binomial experiment with success rate equal to τ . Following Kupiec (1995) and Christoffersen (1998), we do a formal likelihood ratio (LR) test per quantile to determine whether the binomial parameter is significantly different from τ_0 . Formally, for a given value of τ_0 , we test the null hypothesis $H_0 : \tau = \tau_0$ against the alternative hypothesis $H_1 : \tau \neq \tau_0$. Define¹⁰ the exception indicator as

$$I = \begin{cases} 1 & \text{if } C < \hat{Q}_C(\tau_0 | X); \\ 0 & \text{if } C \geq \hat{Q}_C(\tau_0 | X), \end{cases} \quad (5)$$

where $Q_C(\tau | X)$ denotes the τ th conditional quantile of trading costs C given trade

¹⁰Clearly I depends on τ_0 , but we omit any subscripts for simplicity of notation.

characteristics and market circumstances X . Let $z = \sum_{i=1}^m I_i$ be the total number of exceptions in the out-of-sample period. The LR test of correct unconditional coverage has the form

$$LR = 2\{\log(\hat{\tau}^z(1 - \hat{\tau})^{m-z}) - \log(\tau_0^z(1 - \tau_0)^{m-z})\}, \quad (6)$$

where $\hat{\tau} = z/m$. The LR statistic is asymptotically χ^2 distributed with one degree of freedom. The above LR test is a test for unconditional coverage of the VaR estimates, since it is simply based on the total number of exceptions over the entire out-of-sample period. We test the null hypothesis of correct unconditional coverage for several values of τ_0 . The upper panel of Table 5 (with the caption ‘QR’) reports the outcomes of the LR tests, both in-sample and out-of-sample (see the columns with the caption ‘LR’). It is no surprise that the in-sample forecasts have virtually perfect unconditional coverage, as this is inherent to the quantile regression model. Therefore, the out-of-sample forecasts are more interesting. Although the null hypothesis of correct unconditional coverage is rejected for some quantiles in the case of buys, the quantile regression model generally performs well. Hence, rather than relying on relatively inaccurate expected costs, using forecasted conditional quantiles seems a better way to screen trades and to filter out expensive ones.

We could also use the classical regression model to forecast the distribution of market impact costs.¹¹ For completeness’ sake, we test for correct coverage in the linear regression model with empirical error term distribution; see the lower panel of Table 5 (captioned ‘LM’).¹² Again it is no surprise that the in-sample forecasts exhibit exact unconditional coverage, since this is inherent in the empirical

¹¹Alternatively, we could compare the quantile regression model to the *heteroskedastic* regression model. However, at some points this model produces negative values of the conditional standard deviation. This is not a surprise, since the model does not restrict the conditional standard deviation to positive values. Instead, it is more practical to work with a different specification for the conditional variance, for instance $\exp(X\gamma)'$ or $(X\gamma)^2$. However, our investigations show that such a specification performs very similarly to the homoskedastic regression model. Therefore, we apply the simplest specification, which is the latter model.

¹²The empirical distribution is based on the (in-sample) observed model residuals and assigns equal probability mass to each observed value.

error term distribution. The out-of-sample forecasts confirm that the traditional regression model has less accurate coverage than the quantile regression model. The linear model with normal error terms performs even worse, but to save space we do not report these results. The results confirm our earlier finding that the linear regression model is rejected in favor of the more general quantile regression model.

For alternative ways to forecast market impact costs, we refer to Bikker et al. (2006).

5 Conclusions

When a relatively small group of expensive equity transactions determines the main part of market impact costs, substantial cost savings can be realized by reducing the trading costs of comparatively few trades. To do so, expensive trades need to be identified as such before actual trading takes place. This requires accurate modeling of the entire cost distribution; in particular of how trade properties and market circumstances influence the trading costs of the most expensive trades.

The traditional regression approach focuses on the average impact of certain trade properties and market circumstances on trading costs rather than on the extreme effects that these determinants occasionally have. This paper uses a different approach and relies on quantile regression to assess the determinants of extreme market impact costs. Using the flexible quantile regression approach, we can explicitly determine the impact of trade characteristics and market conditions on the $100\tau\%$ most expensive trades, where τ can take any value in the range $0 - 1$.

Previous research has established important roles for trade style and variables related to trade difficulty (such as market capitalization and trade size) in explaining market impact costs. Analyzing data on the world-wide equity trades executed by the world's second largest pension fund during the first quarter of 2002, our study shows that extreme values of momentum and high volatility substantially increase the risk of incurring high trading costs. Our analysis also makes clear that agency

trades are considerably more risky in terms of trading costs than similar principal trades. Moreover, trade timing also turns out to be an important risk factor.

In practice, a major motive for estimating market impact models is their use for prediction purposes in trading cost management. Formal statistical tests confirm that the quantile regression model succeeds well in forecasting the cost distribution. Moreover, it outperforms the traditional regression model in terms of forecasting power.

All in all, this paper shows that quantile regression and the forecasts based on this method can make a productive contribution to transaction cost management.

Acknowledgements

The authors are grateful to an anonymous referee, the participants of the research seminars at De Nederlandsche Bank and University of Twente, as well as the participants of the 2006 Dauphine workshop in Paris and the 2006 ESEM in Vienna. The usual disclaimer applies.

References

- Bikker, J.A., L. Spierdijk, and P.J. van der Sluis (2007). Market Impact Costs of Institutional Equity Trades. Forthcoming *Journal of International Money and Finance*.
- Bikker, J.A., L. Spierdijk, R.P.M.M. Hoevenaars, and P.J. van der Sluis (2006). Forecasting Market Impact Costs and Identifying Expensive Trades. DNB Working Paper 95.
- Buchinsky, M. (1998). Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. *Journal of Human Resources* 33, 88-126.
- Chan, L.K.C. and J. Lakonishok (1993). Institutional Trades and Intraday Stock Price Behavior. *Journal of Financial Economics* 33, 173-199.
- Chan, L.K.C. and J. Lakonishok (1995). The Behavior of Stock Prices around Institutional Trades. *Journal of Finance* 50, 1147-1174.
- Chan, L.K.C. and J. Lakonishok (1997). Institutional Equity Trading Costs: NYSE versus Nasdaq. *Journal of Finance* 52, 713-735.
- Chernozhukov, V. (2005). Extremal Quantile Regression. *Annals of Statistics* 33, 806-839.
- Collins, B.M. and F.J. Fabozzi (1991). A Methodology for Measuring Transaction Costs. *Financial Analysts Journal* 47, 27-36.
- Christoffersen, P.F (1998). Evaluating Interval Forecasts. *International Economic Review* 39, 841-862.
- Easley, D. and M. O'Hara (1987). Price, Size and the Information in Security Markets. *Journal of Financial Economics* 16, 69-90.
- Foster, D.F. and S. Viswanathan (1993). Variations in Trading Volume, Return Volatility, and Trading Costs: Evidence on Recent Price Formation Models. *Journal of Finance* 48, 187-211.
- Grinold, R.C. and R.N. Kahn (1999). *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk*. McGraw-Hill Professional Publishing.
- Hallock, K., R. Madalozzo, and C. Reck (2004). Uncovering Heterogeneity in Managerial Pay-Firm Performance Relationships Using Quantile Regression. Mimeo.
- Heckman, J.J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models. *The Annals of Economic and Social Measurement* 5, 475-492.
- Heckman, J.J. (1979). Sample Selection Bias as a Specification Error. *Econometrica* 47, 153-162.

- Holthausen, R., R. Leftwich, and D. Mayers (1987). The Effect of Large Block Transactions on Security Prices: A Cross-Sectional Analysis. *Journal of Financial Economics* 19, 237-268.
- Holthausen, R., R. Leftwich, and D. Mayers (1990). Large-Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects. *Journal of Financial Economics* 26, 71-95.
- Keim, D.B. and A. Madhavan (1996). The Upstairs Market for Large Block Transactions: Analysis and Measurement of Price Effects. *Review of Financial Studies* 9, 1-36.
- Keim, D.B. and A. Madhavan (1997). Transaction Costs and Investment Style: An Inter-Exchange Analysis of Institutional Equity Trades. *Journal of Financial Economics* 46, 265-292.
- Keim, D.B. and A. Madhavan (1998). The Cost of Institutional Equity Trades. *The Financial Analysts Journal* 54, 50-69.
- Khmaladze, E.V. (1981). Martingale Approach in the Theory of Goodness-of-Fit Tests. *Theory of Probability and its Applications* 26, 420-447.
- Koenker, R. and G.W. Bassett (1978). Regression Quantiles. *Econometrica* 46, 33-50.
- Koenker, R. and K.F. Hallock (2001). Quantile Regression. *Journal of Economic Perspectives* 15, 153-156.
- Koenker, R.W. and V. d'Orey (1987). Computing Regression Quantiles. *Applied Statistics* 36, 383-393.
- Koenker, R.W. and V. d'Orey (1994). Computing Regression Quantiles. *Applied Statistics* 43, 410-414.
- Koenker, R.W. and S. Portnoy (1999). Quantile Regression. Unpublished Manuscript. University of Illinois.
- Koenker, R. and Z. Xiao (2002). Inference on the Quantile Regression Process. *Econometrica* 70, 1583-1612.
- Koenker, R. and J. Machado (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association* 94, 1296-1310.
- Koenker, R. (2005). Quantreg: Quantile Regression, R package version 3.90. <http://www.econ.uiuc.edu/~roger/research/rq/rq.html>.
- Korajczyk, R.A. and R.S. Sadka (2004). Are Momentum Profits Robust to Trading Costs? *Journal of Finance* 59, 1039-1082.
- Kraus, A. and Stoll, H. (1972). Price Impacts of Block Trading on the New York

- Stock Exchange. *Journal of Finance* 27, 569-588.
- Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives* 3, 73-84.
- Lesmond, D.A., M.J. Schill and C. Zhou (2004). The Illusory Nature of Momentum Profits. *Journal of Financial Economics* 71, 349-380.
- Madhavan, A. and M. Cheng (1997). In Search of Liquidity: Block Trades in the Upstairs and Downstairs Market. *Review of Financial Studies* 10, 175-203.
- McInish, T. and R. Wood (1992). An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks. *Journal of Finance* 47, 753-764.
- Moreno, D., P. Marco, and I. Olmeda (2005). Risk forecasting models and optimal portfolio selection. *Applied Economics* 37, 1267-1281.
- Powell, J. (1986). Censored Regression Quantiles. *Journal of Econometrics* 32, 143-155.
- R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. ISBN 3-900051-07-0. <http://www.R-project.org>.
- Schittenkopf, C., P. Tiño, and G. Dorffner (2002). The Benefits of Information Reduction for Trading Strategies. *Applied Economics* 34, 917-930.
- Smith, B.F., D.A. Turnbull, and R.W. White (2001). Upstairs Market for Principal and Agency Trades: Analysis of Adverse Information and Price Effects. *Journal of Finance* 56, 1723-1746.
- Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *Journal of Human Resources* 33, 127-169.
- Wagner, W.H. and M. Edwards (1993). Best Execution. *Financial Analysts Journal* 49, 65-71.

Appendix A Quantile regression

This appendix briefly explains the quantile regression approach. Moreover, we also discuss the differences between this approach and traditional regression models.

Given market impact costs (C) and its determinants (contained in the vector X of dimension K), the classical regression model (alternatively referred to as the ‘location’ model) is formulated as

$$C = X\alpha + \sigma\varepsilon, \quad \mathbb{E}(\varepsilon | X) = 0, \quad \mathbb{E}(\varepsilon^2 | X) = 1, \quad (\text{A.1})$$

with α a vector of coefficients of dimension K and $\sigma > 0$. The classical regression model is restrictive, since it only allows covariates to affect the conditional mean (that is, the term $X\alpha$) of the trading costs. This implies that a change in the determinants merely ‘shifts’ the conditional distribution of the dependent variable C . The more flexible heteroskedastic regression model (also called the ‘location-scale’ model) is written as

$$C = X\alpha + (X\gamma)\varepsilon, \quad \mathbb{E}(\varepsilon | X) = 0, \quad \mathbb{E}(\varepsilon^2 | X) = 1. \quad (\text{A.2})$$

Hence, in the location-scale model the variance of the error terms depends on the covariates X through the vector of coefficients γ . This means that the covariates additionally influence the conditional variance of the dependent variable, thus stretching (larger variance) or squeezing (smaller variance) the distribution of trading costs. However, since we expect that market impact costs are affected by its determinants in more complex ways, the usual regression models are presumably too restrictive. Therefore, we consider the more flexible quantile regression approach.

We say that market impact costs incurred on a trade are at the τ th ($0 < \tau < 1$) quantile if they are higher than the proportion τ and lower than the proportion $1 - \tau$ of all trades’ market impact costs. More formally, the τ th quantile of trading costs

C (denoted by $Q_C(\tau)$) is defined as

$$Q_C(\tau) = \inf_c \{c : F_C(c) \geq \tau\}, \quad (\text{A.3})$$

where $F_C(c) = \mathbb{P}(C \leq c)$ denotes the distribution function of C . Definition (A.3) states that the τ th quantile is equal to the smallest value c for which $F_C(c)$ is at least equal to τ . Since we observe response variables (market impact costs) in combination with covariates, we are actually interested in conditional quantiles rather than unconditional ones. For instance, we may want to know what the trading costs will be given a particular trade size and a certain level of stock price volatility. The τ th conditional quantile of C given $X = x$ (denoted by $Q_C(\tau | x)$) is defined in a very similar way:

$$Q_C(\tau | x) = \inf_c \{c : F_{C|X}(c | x) \geq \tau\}, \quad (\text{A.4})$$

where $F_{C|X}(c | x) = \mathbb{P}(C \leq c | X = x)$ denotes the conditional distribution function of C given $X = x$. Just as the classical linear regression models are based on the assumption that the conditional expectation of the dependent variable C given the covariates X is equal to $X\alpha$, the quantile regression model assumes that the τ th conditional quantile of C given the covariates X equals $X\beta_\tau$. That is, we can formulate the quantile regression model in a similar manner as the usual regression model:

$$C = X\beta_\tau + \varepsilon, \quad Q_\varepsilon(\tau | X) = 0. \quad (\text{A.5})$$

In specification (A.5), the partial derivative of C with respect to one of the regressors (say X_j) equals the j th element of the vector β_τ and represents the change in the τ th conditional quantile due to a (ceteris paribus) one-unit change in X_j .

We now turn to the relation between quantile regression and traditional regression models.¹³ Assuming that the regression model contains an intercept, we write $X =$

¹³We notice that it would be clearly wrong to approximate the quantile regression model by dividing the dependent variable into subsets according to its unconditional distribution and by subsequently applying ordinary least squares to these subsets. The example of Hallock, Madalozzo,

[1 Z], $\alpha = (\alpha_0, \alpha_1)$, $\gamma = (\gamma_0, \gamma_1)$, and $\beta_\tau = (\beta_{\tau,0}, \beta_{\tau,1})$. Applying definition (A.4) to the conditional quantiles in the classical linear regression model, we find that the conditional quantiles are given by $Q_C(\tau | X) = X\alpha + \sigma F_\varepsilon^{-1}(\tau) = \alpha_0 + Z\alpha_1 + \sigma F_\varepsilon^{-1}(\tau)$, where F_ε denotes the distribution function of ε . Similarly, in the location-scale model, we find $Q_C(\tau | X) = X\alpha + X\gamma F_\varepsilon^{-1}(\tau) = \alpha_0 + Z\alpha_1 + (\gamma_0 + Z\gamma_1)F_\varepsilon^{-1}(\tau)$. Thus, the location model imposes certain restrictions on the coefficients of the conditional quantiles in the quantile regression model, namely $\beta_{\tau,0} = \alpha_0 + \sigma F_\varepsilon^{-1}(\tau)$ and $\beta_{\tau,1} = \alpha_1$, whereas the location-scale model imposes $\beta_{\tau,0} = \alpha_0 + \gamma_0 F_\varepsilon^{-1}(\tau)$ and $\beta_{\tau,1} = \alpha_1 + \gamma_1 F_\varepsilon^{-1}(\tau)$.

Both the location and the location-scale model are special cases of the quantile regression approach. The crucial difference between the classical and the quantile regression model is that in the former the covariates affect all quantiles in a similar way, whereas in the latter model the impact of the explanatory variable on the τ th quantile depends on τ . Although the location-scale model allows the impact of the covariates to vary over the quantiles as well, it remains more restrictive than the quantile regression model as it imposes certain restrictions on β_τ .

Appendix B Forecast error measures

Given a sample of observations C_1, \dots, C_n and corresponding forecasts $\hat{C}_1, \dots, \hat{C}_n$, the mean absolute percentage error (MAPE) is defined as

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{C}_i - C_i}{C_i} \right|. \quad (\text{B.1})$$

Furthermore, the mean squared error (MSE) is calculated as

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{C}_i - C_i)^2. \quad (\text{B.2})$$

and Reck (2004) demonstrates that such truncation of the dependent variable can lead to erroneous conclusions, due to the selection bias of Heckman (1979).

The bias, variance, and covariance proportions of the MSE are given by

$$BP = \frac{(\bar{\hat{C}} - \bar{C})^2}{\sum_{i=1}^n (\hat{C}_i - C_i)^2/n}, \quad VP = \frac{(s_{\hat{C}} - s_C)^2}{\sum_{i=1}^n (\hat{C}_i - C_i)^2/n}, \quad CP = \frac{2(1 - \hat{\rho})s_{\hat{C}}s_C}{\sum_{i=1}^n (\hat{C}_i - C_i)^2/n},$$

where $\bar{C}, \bar{\hat{C}}, s_C, s_{\hat{C}}$ are the sample means and variances of C_1, \dots, C_n and $\hat{C}_1, \dots, \hat{C}_n$, respectively. The sample correlation between the series of observed values and forecasts is denoted by $\hat{\rho}$. Finally, Theil's inequality coefficient is obtained as

$$U = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{C}_i - C_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \hat{C}_i^2 + \frac{1}{n} \sum_{i=1}^n C_i^2}}. \quad (\text{B.3})$$

| variable | definition |
|------------------|---|
| momentumperc | 5-day volume-weighted average return prior to trading (in %) |
| volatility | logarithm of 30-day individual volatility prior to trading (in %) |
| tradesize | square root of trade size relative to 30-day average daily volume prior to trading (in %) |
| marketcap | logarithm of market capitalization 3 months prior to trading (in billion Euro) |
| adv | logarithm of 30-day average daily trading volume of stock (in shares) |
| exprice | logarithm of execution price of stock (in Euro) |
| agency singledum | 0/1 variable for agency/single (1) or principal (0) trades |
| growthdum | 0/1 variable for growth stocks |
| quantdum | 0/1 variable for trades executed by quantitative (1) or fundamental (0) fund |
| preopendum | 0/1 variable for trades sent to broker during pre-opening of the market |
| morningdum | 0/1 variable for trades sent to broker during in the morning (after pre-opening) |
| middaydum | 0/1 variable for trades sent to broker during in the afternoon |
| (Mondaydum) | 0/1 variable for trades executed on Monday |
| (Tuesdaydum) | 0/1 variable for trades executed on Tuesday |
| Wednesdaydum | 0/1 variable for trades executed on Wednesday |
| Thursdaydum | 0/1 variable for trades executed on Thursday |
| Fridaydum | 0/1 variable for trades executed on Friday |
| earlymonthdum | 0/1 variable for trades executed at the beginning of the month |
| Jandum | 0/1 variable for trades executed in January |
| Febdum | 0/1 variable for trades executed in February |
| (Marchdum) | 0/1 variable for trades executed in March |
| tradedur | logarithm of the time elapsed between the moment that trade was passed to the broker and trade execution |
| NYSEdum | 0/1 variable for trades executed on NYSE |
| Nasdaqdum | 0/1 variable for trades executed on Nasdaq |
| Torontodum | 0/1 variable for trades executed on Toronto Stock Exchange |
| Londondum | 0/1 variable for trades executed on London Stock Exchange |
| Tokyodum | 0/1 variable for trades executed on Tokyo Stock Exchange |
| upstairsdum | 0/1 variable for trades executed on exchange with upstairs market |
| (dealerdum) | 0/1 variable for trades executed on exchange with dealer market |
| LOBdum | 0/1 variable for trades executed on exchange with electronic limit order book |
| (floordum) | 0/1 variable for trades executed on exchange with trading floor |
| (hybriddum) | 0/1 variable for trades executed on exchange with hybrid market (LOB+dealers) |
| mcapdum | logarithm of domestic market capitalization of the exchange on which the stock was traded (in billion Euro) |
| consumerdiscrdum | 0/1 variable for stocks in consumer discretionary sector |
| consumerstdum | 0/1 variable for stocks in consumer staples sector |
| energydum | 0/1 variable for stocks in energy sector |
| finservdum | 0/1 variable for stocks in financial services sector |
| healthdum | 0/1 variable for stocks in health sector |
| ITdum | 0/1 variable for stocks in IT sector |
| materdum | 0/1 variable for stocks in materials sector |
| telecomdum | 0/1 variable for stocks in telecommunications sector |
| utilitiesdum | 0/1 variable for stocks in utilities sector |
| (industrydum) | 0/1 variable for stocks in industry sector |

Table 1: Potential determinants of market impact costs and their definitions

The dummy variables in parentheses have not been included in the estimated models to avoid exact collinearity, but are included in the table for completeness. Since there are virtually no trades on Monday during the sample period, we exclude both the Monday and the Tuesday dummy. The sectors correspond to the Global Industry Classification Standard (GICS).

| | BUYS | | SELLS | |
|----------------|-------------|------------|--------------|------------|
| | weighted | unweighted | weighted | unweighted |
| mean | 19.6 | 16.5 | 29.7 | -0.5 |
| st.dev. mean | 5.7 | 119.5 | 6.5 | 144.2 |
| median | 0.2 | 7.9 | 0.0 | 2.2 |
| 0.5% quantile | -942.7 | -391.5 | -950.7 | -487.6 |
| 5% quantile | -133.1 | -163.0 | -157.3 | -245.5 |
| 95% quantile | 241.6 | 201.9 | 357.5 | 210.7 |
| 99.5% quantile | 1329.5 | 501.6 | 1426.2 | 557.3 |

Table 2: Sample statistics of market impact costs (in bp)

This table presents both principal-weighted and unweighted market impact costs statistics. The weighted statistics are obtained by weighting each observation by the Euro value of the trade, so that larger trades contribute more to, for instance, the average market impact than smaller ones.

| BUYS | | | | SELLS | | | |
|-----------------|------------------|---------------|-------------------|-----------------|------------------|---------------|-------------------|
| | <i>left tail</i> | <i>median</i> | <i>right tail</i> | | <i>left tail</i> | <i>median</i> | <i>right tail</i> |
| momentumperc | 0.0422 | 0.0024 | 0.0145 | momentumperc | 0.0054 | 0.0066 | 0.0591 |
| volatility | 0.0402 | 0.0001 | 0.0159 | volatility | 0.0369 | 0.0004 | 0.0277 |
| marketcap | 0.0128 | 0.0010 | 0.0010 | marketcap | 0.0019 | 0.0028 | 0.0201 |
| tradesize | 0.0022 | 0.0019 | 0.0014 | tradesize | 0.0041 | 0.0025 | 0.0030 |
| agencysingledum | 0.0067 | 0.0242 | 0.0485 | agencysingledum | 0.0215 | 0.0038 | 0.0395 |
| quantdum | 0.0130 | 0.0152 | 0.0012 | quantdum | 0.0014 | 0.0104 | 0.0129 |
| preopendum | 0.0034 | 0.0247 | 0.0237 | preopendum | 0.0191 | 0.0011 | 0.0006 |
| morningdum | 0.0150 | 0.0226 | 0.0106 | morningdum | 0.0233 | 0.0092 | 0.0008 |
| earlymonthdum | 0.0021 | 0.0014 | 0.0010 | earlymonthdum | 0.0194 | 0.0049 | 0.0032 |
| Jandum | 0.0010 | 0.0012 | 0.0338 | Jandum | 0.0191 | 0.0005 | 0.0061 |
| NYSEdum | 0.0040 | 0.0031 | 0.0160 | NYSEdum | 0.0290 | 0.0113 | 0.0025 |
| Wednesdaydum | 0.0179 | 0.0338 | 0.0233 | Wednesdaydum | 0.0023 | 0.0187 | 0.0053 |
| Thursdaydum | 0.0292 | 0.0357 | 0.0034 | Thursdaydum | 0.0016 | 0.0154 | 0.0121 |
| Fridaydum | 0.0218 | 0.0353 | 0.0147 | Fridaydum | 0.0234 | 0.0250 | 0.0020 |
| tradedur | 0.0090 | 0.0067 | 0.0085 | Nasdaqdum | 0.0211 | 0.0327 | 0.0258 |
| Nasdaqdum | 0.0052 | 0.0041 | 0.0098 | | | | |

Table 3: Squared partial correlations

Based on the quantile regression model, Tables 3 displays the values of the squared partial correlation (SPC) for each explanatory variable in the left and right tail and the center of the cost distribution. The most important variables according to the SPC are in bold face.

| | BUYS | | SELLS | |
|--------------------------------|------------------|----------------------|------------------|----------------------|
| | <i>in-sample</i> | <i>out-of-sample</i> | <i>in-sample</i> | <i>out-of-sample</i> |
| Theil's U | 0.61 | 0.64 | 0.57 | 0.69 |
| mean absolute percentage error | 4.1 | 13.8 | 3.3 | 4.7 |
| mean squared error | 12,874 | 9,960 | 17,396 | 13,004 |
| bias part (%) | 0.0 | 6.7 | 0.0 | 15.1 |
| variance part (%) | 39.0 | 19.9 | 33.1 | 28.6 |
| covariance part (%) | 61.1 | 73.6 | 67.0 | 56.6 |
| naive hit ratio (%) | 59.2 | 52.7 | 50.5 | 59.1 |
| hit ratio (%) | 67.7 | 54.0 | 66.3 | 48.4 |
| correlation to realized costs | 0.44 | 0.19 | 0.50 | 0.28 |

Table 4: Error measures for in-sample and out-of-sample forecasts

Table 4 displays various error measures for the in-sample and out-of-sample forecasts of market impact costs based on the ordinary least squares estimator.

| BUYS | | | SELLS | | | | | |
|--------------|--------------------|-----|----------------------|-------------|--------------------|-----|----------------------|-------------|
| QR | | | | | | | | |
| | <i>in-sample</i> | | <i>out-of-sample</i> | | <i>in-sample</i> | | <i>out-of-sample</i> | |
| τ_0 (%) | $\hat{\tau}_0$ (%) | LR | $\hat{\tau}_0$ (%) | LR | $\hat{\tau}_0$ (%) | LR | $\hat{\tau}_0$ (%) | LR |
| 5 | 5.2 | 0.0 | 2.0 | 4.9 | 5.5 | 0.3 | 4.1 | 0.3 |
| 10 | 10.5 | 0.2 | 6.6 | 2.9 | 9.5 | 0.2 | 8.8 | 0.3 |
| 15 | 14.7 | 0.1 | 9.8 | 4.6 | 15.2 | 0.0 | 14.8 | 0.0 |
| 20 | 20.3 | 0.0 | 14.7 | 3.8 | 20.2 | 0.0 | 18.9 | 0.1 |
| 25 | 25.1 | 0.0 | 21.7 | 1.2 | 25.3 | 0.0 | 30.3 | 2.4 |
| 30 | 30.0 | 0.0 | 40.0 | 9.0 | 29.6 | 0.0 | 33.9 | 1.2 |
| 35 | 35.5 | 0.0 | 48.1 | 14.4 | 34.5 | 0.0 | 38.9 | 1.1 |
| 40 | 40.0 | 0.0 | 36.3 | 1.1 | 39.5 | 0.0 | 41.2 | 0.1 |
| 45 | 45.2 | 0.0 | 29.5 | 19.9 | 44.7 | 0.0 | 43.8 | 0.1 |
| 50 | 50.2 | 0.0 | 30.2 | 32.0 | 49.8 | 0.0 | 50.8 | 0.0 |
| 55 | 55.5 | 0.0 | 34.4 | 34.2 | 55.1 | 0.0 | 54.4 | 0.0 |
| 60 | 60.5 | 0.0 | 43.3 | 22.4 | 59.9 | 0.0 | 65.3 | 2.0 |
| 65 | 65.4 | 0.0 | 63.0 | 0.3 | 65.6 | 0.0 | 68.7 | 1.0 |
| 70 | 70.0 | 0.0 | 73.3 | 1.1 | 70.6 | 0.0 | 73.3 | 0.9 |
| 75 | 74.6 | 0.0 | 84.9 | 11.6 | 75.5 | 0.0 | 75.9 | 0.1 |
| 80 | 80.1 | 0.0 | 88.2 | 9.4 | 80.1 | 0.0 | 83.7 | 1.5 |
| 85 | 84.5 | 0.1 | 90.2 | 4.6 | 85.1 | 0.0 | 88.9 | 2.1 |
| 90 | 90.0 | 0.0 | 92.1 | 1.1 | 89.5 | 0.2 | 90.2 | 0.0 |
| 95 | 95.1 | 0.0 | 97.2 | 2.3 | 94.6 | 0.2 | 95.6 | 0.1 |
| LM | | | | | | | | |
| 5 | 5.0 | 0.0 | 7.4 | 2.2 | 5.0 | 0.0 | 7.8 | 2.3 |
| 10 | 10.0 | 0.0 | 17.5 | 10.4 | 10.1 | 0.0 | 17.1 | 7.9 |
| 15 | 15.0 | 0.0 | 23.2 | 9.2 | 15.0 | 0.0 | 28.0 | 18.5 |
| 20 | 20.0 | 0.0 | 33.7 | 20.4 | 20.0 | 0.0 | 36.3 | 23.8 |
| 25 | 25.0 | 0.0 | 42.5 | 28.7 | 25.1 | 0.0 | 43.5 | 27.2 |
| 30 | 30.0 | 0.0 | 55.6 | 55.8 | 30.0 | 0.0 | 50.8 | 31.5 |
| 35 | 35.0 | 0.0 | 65.0 | 73.7 | 35.0 | 0.0 | 57.5 | 35.2 |
| 40 | 40.0 | 0.0 | 71.3 | 79.8 | 40.0 | 0.0 | 61.7 | 31.9 |
| 45 | 45.0 | 0.0 | 75.5 | 76.4 | 45.0 | 0.0 | 66.8 | 32.4 |
| 50 | 50.0 | 0.0 | 79.6 | 74.6 | 50.0 | 0.0 | 71.0 | 30.5 |
| 55 | 55.0 | 0.0 | 83.8 | 74.4 | 55.0 | 0.0 | 74.9 | 28.3 |
| 60 | 60.0 | 0.0 | 85.8 | 63.3 | 60.0 | 0.0 | 76.7 | 20.9 |
| 65 | 65.0 | 0.0 | 88.4 | 57.1 | 65.0 | 0.0 | 78.5 | 14.5 |
| 70 | 70.0 | 0.0 | 90.2 | 47.0 | 70.0 | 0.0 | 81.1 | 10.7 |
| 75 | 75.0 | 0.0 | 91.0 | 33.5 | 74.9 | 0.0 | 82.6 | 5.7 |
| 80 | 80.0 | 0.0 | 93.0 | 26.4 | 80.0 | 0.0 | 85.8 | 3.8 |
| 85 | 85.0 | 0.0 | 94.3 | 17.0 | 85.0 | 0.0 | 87.6 | 0.9 |
| 90 | 90.0 | 0.0 | 96.3 | 11.2 | 89.9 | 0.0 | 89.9 | 0.0 |
| 95 | 95.0 | 0.0 | 97.6 | 3.4 | 95.0 | 0.0 | 93.8 | 0.5 |

Table 5: Outcomes of likelihood ratio test for unconditional coverage

This table displays the outcomes of the likelihood ratio test for unconditional coverage for the in-sample and the out-of-sample period. A test statistic LR in boldface indicates rejection of the null hypothesis $H_0 : \tau = \tau_0$ at a 5% level. The outcomes of the likelihood ratio test are given for both the quantile regression model ('QR') and the linear model with empirical error term distribution ('LM').

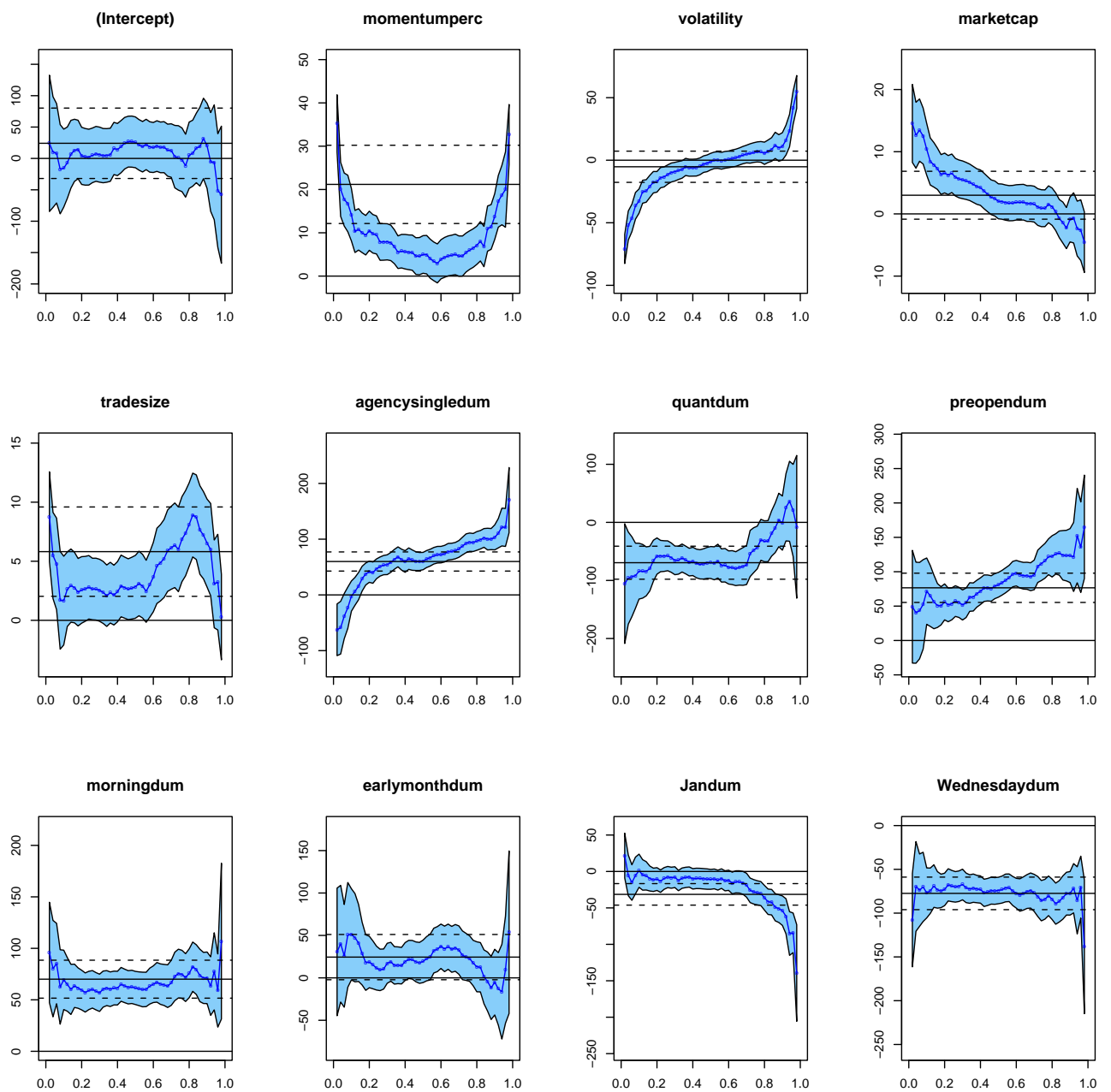


Figure 1: Estimated quantile regression coefficients with confidence bounds (buys)

Figures 1-4 display the impact of the trade characteristics on the distribution of the market impact costs as a function of the quantile (solid curve) and 95% confidence bounds (light-shaded area). The estimation results are based on the Barrodale and Roberts algorithm implemented in the Roger Koenker's Quantreg package for R. Additionally, ordinary least squares coefficients (solid black line) and corresponding 95% confidence bounds (dashed black lines) obtained from the classical linear regression model are plotted to facilitate visual comparison of the quantile regression with the traditional regression approach.

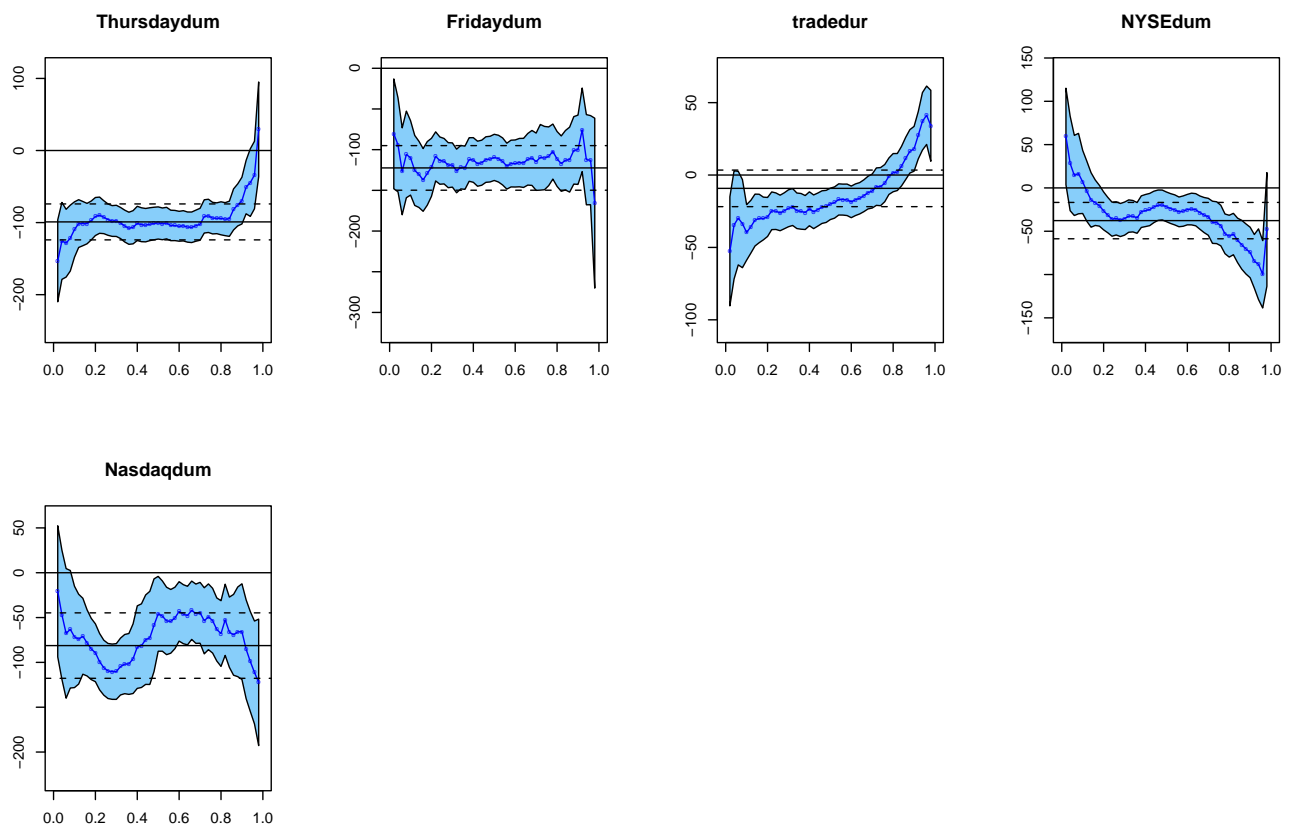


Figure 2: Estimated quantile regression coefficients with confidence bounds (buys, continued)

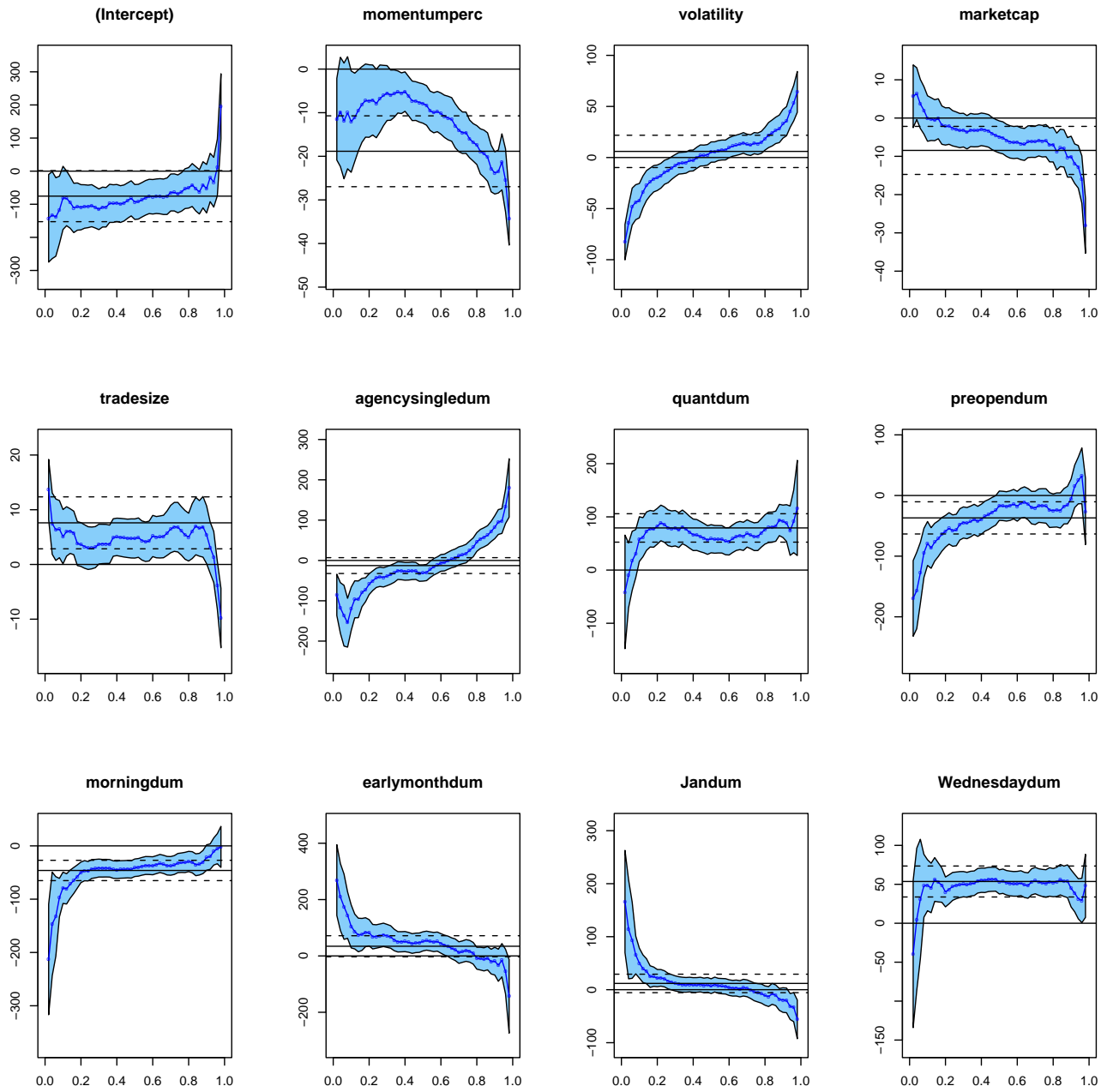


Figure 3: Estimated quantile regression coefficients with confidence bounds (sells)

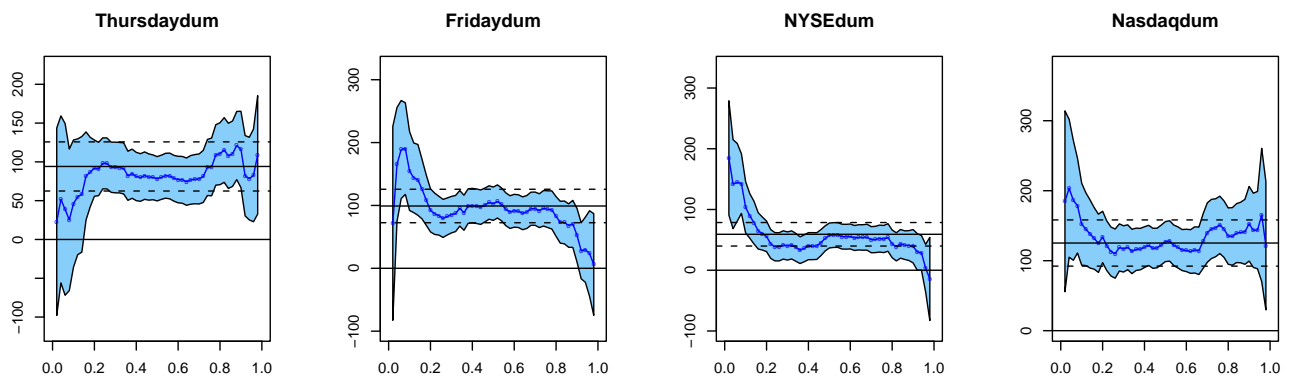


Figure 4: Estimated quantile regression coefficients with confidence bounds (sells, continued)

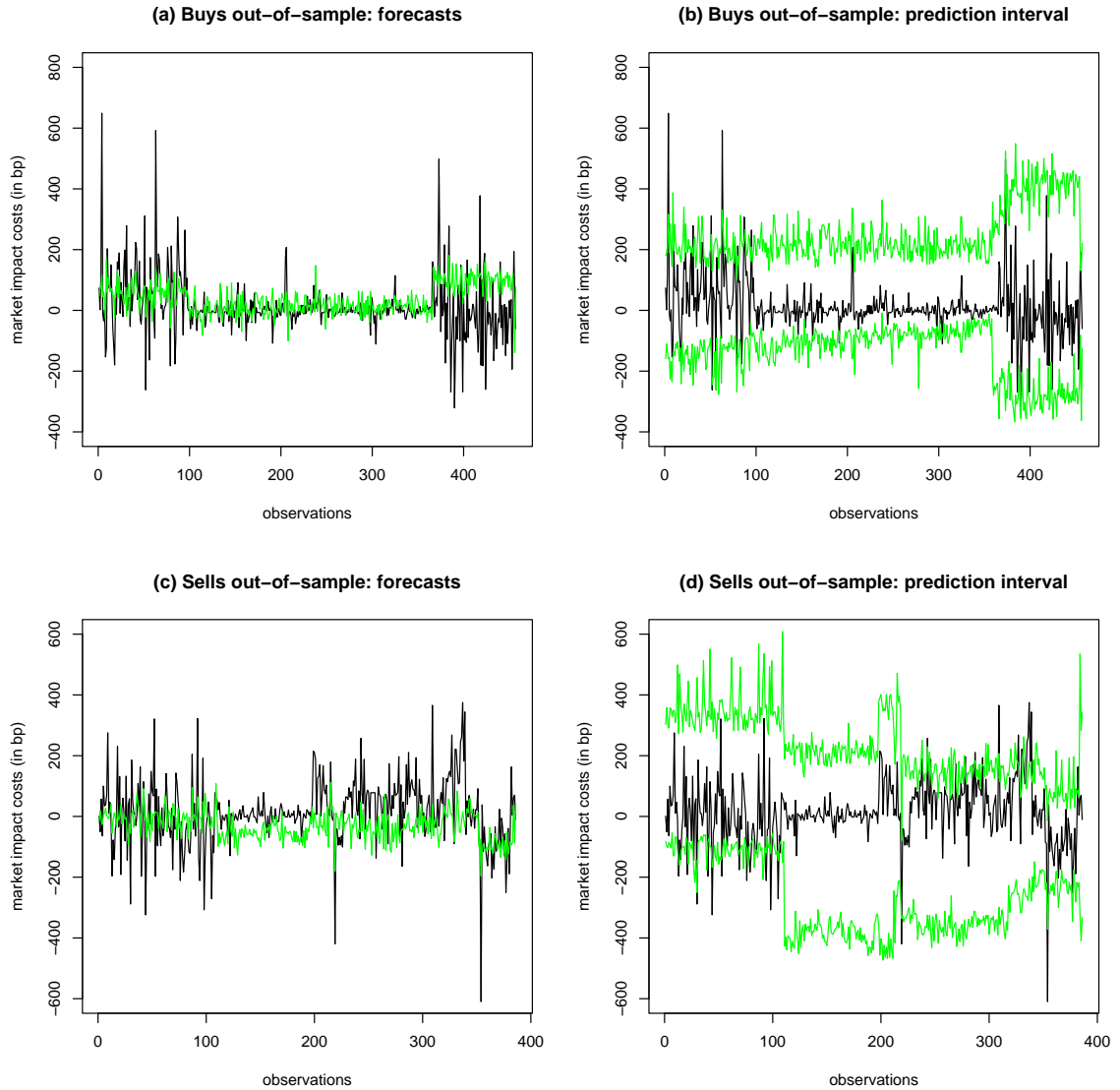


Figure 5: Forecasted and realized market impact costs

Figures (a) and (c) display forecasted (light-colored lines) and realized (dark-colored lines) market impact costs for buys and sells during the out-of-sample period. The forecasts are based on the linear regression model. The light-colored lines in Figures (b) and (d) depict the upper and lower bounds of the 90% prediction interval obtained from the quantile regression model; the dark-colored lines represent realized market impact costs.