

# VU Research Portal

## Riskfree Rate Dynamics

van der Wel, M.

2008

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

van der Wel, M. (2008). *Riskfree Rate Dynamics: Information, Trading, and State Space Modeling*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

Riskfree Rate Dynamics:  
Information, Trading, and State Space Modeling

ISBN 978 90 5170 769 4

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 433 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which have already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

Riskfree Rate Dynamics:  
Information, Trading, and State Space Modeling

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan  
de Vrije Universiteit Amsterdam,  
op gezag van de rector magnificus  
prof.dr. L.M. Bouter,  
in het openbaar te verdedigen  
ten overstaan van de promotiecommissie  
van de faculteit der Economische Wetenschappen en Bedrijfskunde  
op maandag 1 december 2008 om 10.45 uur  
in de aula van de universiteit,  
De Boelelaan 1105

door

Michel van der Wel

geboren te Amsterdam

promotoren: prof.dr. A. Lucas  
prof.dr. S.J. Koopman  
copromotor: dr. A.J. Menkveld

## Acknowledgments

This dissertation is the result of four tough but fascinating years. I am happy to have the opportunity to thank my supervisors for their help and guidance during this time. Thank you Siem Jan Koopman, for interesting me in academics during my econometrics bachelor and master study. We have been working together now for almost 6 years, I look forward to continuing this. Thank you Albert Menkveld, for sharing with me your passion for financial markets. You have invested a great deal of time in me, I learned so much from every interaction we had. Thank you Andre Lucas, I enjoyed our meetings after which I often had a fresh angle to our results and knew which way to go with the research. In addition to my supervisors I owe a great deal of thanks to Asani Sarkar at the Federal Reserve Bank of New York. I enjoyed our conversations, whether they were by mail, phone or in person in either Amsterdam or New York.

Secondly I thank all the researchers I have interacted with over the past years. I am fortunate enough to have presented my work during 24 occasions at 15 cities in 8 countries. During these presentations I appreciated the many comments and challenging but insightful questions. In addition I benefited from interactions with many of the researchers and staff of the VU's Finance and Econometrics groups. I enjoyed all the talks that improved my work, but also just the regular coffee-machine talk about soccer, politics, holidays and other small-talk. In particular I like to thank Charles Bos, Riccardo Calcagno, Drew Creal, Milada Gadourek, Roman Kräussl, Michael Massmann, Marius Ooms and Arjen Siegman.

The four years would not be as interesting as they were without my fellow PhD students. At the Tinbergen Institute I enjoyed the time spent with Eddy Bekkers, Cui Jiajia, Sumedha Gupta, Huang Jian, Lin Tse-Chun, Marta Lopez, Linda Midgley, Tseng Yin-Yen and Ronald Wolthoff. At the VU I was happy to be around Mahmoud Botshekan, Irma Hindrayanto, Borus Jungbacker, Bahar Kaynar, Sander Konijn, Bernd Schwaab, Wang Ting, Brian Wong and Yuan Taoying. Also many thanks to the people I met during my econometrics study. I am fortunate enough to still be in touch with a lot of you, in particular Lianne Bogaard, Bart Bonekamp, Sybren van Boven, Aukje van den Hoeven, Maarten van Kampen, Ofer Tzvi, Ricardo Westra and Teun Zijp.

The final and most important thanks go to my parents, sister, girlfriend (or better: soon-to-be wife) and family. Pa, ma en Ingrid, enorm bedankt voor de steun en begrip die jullie me over de afgelopen jaren hebben gegeven. Zonder jullie was het allemaal nooit gelukt. Pan Lei, meeting you was the best experience during the last four years. Thanks for all your support, I am so happy to have a loving partner that is experiencing the same difficulties. I dedicate this dissertation to you, my mother, father and sister.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background of U.S. Government Bond Market . . . . .	2
1.2	Datasets Analyzed and Relation Between Them . . . . .	3
1.3	Trading U.S. Government Bonds . . . . .	5
1.3.1	Market Microstructure . . . . .	5
1.3.2	The Importance of Customer Flow . . . . .	7
1.3.3	Market Makers . . . . .	8
1.4	Yields of U.S. Government Bonds . . . . .	10
1.4.1	Modeling the Yield Curve . . . . .	10
1.4.2	Extensions to the Nelson-Siegel (1987) Model . . . . .	11
<b>2</b>	<b>Customer Flow, Intermediaries, and the Discovery of the Equilibrium Riskfree Rate</b>	<b>13</b>
2.1	Introduction . . . . .	13
2.2	Background, Data, and Summary Statistics . . . . .	18
2.2.1	Background . . . . .	18
2.2.2	Data . . . . .	18
2.2.3	Summary Statistics . . . . .	21
2.3	Customer Order Informativeness . . . . .	27
2.3.1	Five-Minute Price Change Regressions on Customer Flow . . . . .	27
2.3.2	An Alternative Interpretation of the Regression Coefficient . . . . .	31
2.4	Intermediary's Own-Account Trading . . . . .	32
2.4.1	Is Direct Access to Customer Flow Profitable? . . . . .	32
2.4.2	The Alternative Explanation: Superior Trading Skills . . . . .	36
2.4.3	Do Profits Increase with the Level of Customer Flow Access? . . . . .	38
2.5	Who Pays for the Dual's Increased Profit? . . . . .	40
2.5.1	Profitability of Dual- and Broker-Intermediated Customer Orders . . . . .	41
2.6	Conclusion . . . . .	43
	Appendix 2A: Flow-based Speculation . . . . .	44



Appendix 2B: Additional Tables . . . . .	46
<b>3 Are Market Makers Liquidity Suppliers?</b>	<b>49</b>
3.1 Introduction . . . . .	49
3.2 Related Literature . . . . .	54
3.2.1 Market Making and Inventory . . . . .	54
3.2.2 Signing Trades in the Absence of Quotes . . . . .	55
3.3 Data and Institutional Background . . . . .	56
3.4 Signing Trades in the Absence of Quotes . . . . .	60
3.4.1 Methodology . . . . .	60
3.4.2 Simulation Study and Results . . . . .	62
3.5 Empirical Results . . . . .	66
3.5.1 Initiated Trades of the Market Makers . . . . .	66
3.5.2 Market Maker Inventory and Macro News . . . . .	70
3.5.3 Inventory Increasing Trades . . . . .	70
3.5.4 Inventory Increasing Trades and Trading Profits . . . . .	73
3.5.5 Bid-Ask Spread of Market Makers . . . . .	78
3.6 Conclusion . . . . .	80
Appendix 3A: Recursions of Filter . . . . .	82
<b>4 Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters</b>	<b>89</b>
4.1 Introduction . . . . .	89
4.2 The Dynamic Nelson-Siegel Model . . . . .	92
4.2.1 The Nelson-Siegel Model . . . . .	92
4.2.2 The Dynamics of the Latent Factors . . . . .	94
4.2.3 Estimation Based on the Kalman Filter . . . . .	94
4.3 DNS Model with Time-Varying Parameters . . . . .	95
4.3.1 Time-Varying Loading Parameter . . . . .	96
4.3.2 Time-Varying Volatility . . . . .	97
4.3.3 Estimation Based on the Extended Kalman Filter . . . . .	99
4.4 Data and Empirical Findings . . . . .	101
4.4.1 DNS: Baseline Dynamic Nelson-Siegel Model . . . . .	103
4.4.2 DNS–TVL: Time-Varying Factor Loadings . . . . .	105
4.4.3 DNS–GARCH: Time-Varying Volatility . . . . .	109
4.4.4 DNS–TVL–GARCH: Time-Varying Loadings and Volatility . . . . .	111
4.4.5 Robustness of Empirical Results . . . . .	116
4.5 An Illustration for an Unbalanced Dataset . . . . .	118

4.6	Conclusion . . . . .	120
<b>5</b>	<b>Conclusion</b>	<b>121</b>
5.1	Trading U.S. Government Bonds . . . . .	121
5.2	Yields of U.S. Government Bonds . . . . .	122
5.3	Policy Implications . . . . .	123
5.4	Directions for Further Research . . . . .	123
	<b>Samenvatting (Summary in Dutch)</b>	<b>125</b>
	<b>Bibliography</b>	<b>132</b>



# Chapter 1

## Introduction

The bond market is the largest financial market. The total outstanding volume is more than \$77 trillion, compared to \$60 trillion outstanding equity market volume. The United States accounts for 38% of the global bond market value, with \$6.5 trillion coming from government issued bonds.<sup>1</sup> In this dissertation I look at various aspects of government bonds (also known as *treasuries*) issued by the United States.

An interesting aspect of looking at U.S. treasuries, besides the huge size of the market, is that they serve as a benchmark for many other traded securities. As they are issued by the government the risk of default is negligible. The probability that the United States does not meet its payment obligations is extremely small, as the government controls the supply of money. Due to this property the return on government bonds is commonly referred to as the *riskfree rate*.<sup>2</sup> The yields earned on U.S. treasuries serve as benchmark for all other dollar denominated bonds that are subject to default risk, such as those issued by corporations. Moreover, the results in this dissertation for the U.S. government bond market also provide insights relevant for (European) government bond markets in particular, and financial markets in general.

This dissertation is organized as follows. In the first part, consisting of two chapters, I look at various aspects of trading U.S. government bonds. In the second part of the dissertation, containing one chapter, I look at a popular model to describe the yields offered on U.S. treasuries and provide several extensions. First, in the current chapter I provide a general introduction. I start with giving the institutional setting of the market

---

<sup>1</sup>The bond market volume is from the Bank for International Settlements (BIS) Quarterly Review of March 2008. The data used is the most recent complete available data (September 2007) from the statistical appendix. The equity market volume comes from the World Federation of Exchanges, and is also for September 2007.

<sup>2</sup>This does not mean government bonds are completely free of any risk. Due to inflation the real amount of the coupons and face value changes over time, and there is risk due to changes in interest rates if the bond is held for a time shorter than the remaining maturity. Strictly speaking only government bonds expiring in very short-term can be considered riskfree.

for U.S. government bonds and discussing the datasets that I use to analyze this market. I finish this chapter with describing where each of the remaining chapters fits in the broad (financial) economics literature and provide a brief introduction to each of the chapters in this dissertation.

## 1.1 Background of U.S. Government Bond Market

Treasuries issued by the U.S. government trade with various maturities, typically classified in three groups. Treasuries expiring within a year are commonly referred to as *bills*, treasuries expiring between one and ten years are referred to as *notes* and treasuries expiring in ten years or more are referred to as *bonds*.<sup>3</sup> The U.S. government issues each of the treasuries at regular times in a year through an auction. For example, U.S. government bonds with a maturity of 30 years are issued quarterly. Through the issue of the bonds the government collects the liquidity it needs for the public finances. The auctioning of the treasuries is often referred to as the *primary market*.<sup>4</sup>

After the bonds are auctioned they enter what is called the *secondary market*. In this market they trade both over-the-counter (also referred to as the *spot market*) and in futures markets.<sup>5</sup> As described in Fleming (1997), the spot market evolves around a large group of 1,700 brokers and dealers, with the majority of trading volume coming from 39 primary dealers. These primary dealers are firms with which the Federal Reserve Bank of New York interacts directly. They trade with customers (such as nonprimary dealers, banks, insurance companies, pension funds, and mutual funds) and with each other through interdealer trades. Interdealer trades take place either directly or through 6 interdealer brokers. In his 1994 sample period Fleming (1997) calculates that \$67 billion out of \$125.5 billion daily trading volume on the spot market is from trades between a primary dealer and a customer, and \$58.5 billion is from trades between primary dealers. Of this \$58.5 billion the great majority, \$53.5 billion, is from interdealer trades through interdealer brokers. Mizrach and Neely (2006) document the recent transition from the above described voice-based trading through interdealer brokers to electronic trading. In the time-span of two years two significant Electronic Communication Networks (ECNs) were introduced: in 1999 Cantor Fitzgerald introduced eSpeed and in 2000 a large consortium of Wall Street firms introduced BrokerTec. Trading nowadays is almost completely electronic, with the market split about 39% and 61% for eSpeed and BrokerTec respec-

---

<sup>3</sup>However, throughout this dissertation I often use the label ‘bonds’ to refer to treasuries in general.

<sup>4</sup>For a recent description of the U.S. treasury auction process, see Garbade and Ingber (2005).

<sup>5</sup>At times the treasuries also trade on various other exchanges, such as the New York Stock Exchange and the American Stock Exchange. But, as Fleming (1997) points out, volume on these markets is negligible compared to that of the other markets.

tively.

Besides the spot market significant volume also trades in the futures market. Futures contracts with as underlying treasury notes and bonds trade on the Chicago Board of Trade (CBOT), contracts with as underlying treasury bills trade on the Chicago Mercantile Exchange (CME).<sup>6</sup> On CBOT it is for example possible to trade in 30 Year U.S. Treasury Bonds Futures and the 10, 5 and 2 Year U.S. Treasury Notes Futures contracts. The deliverable often allows for a slight deviation from the maturity in its name. For example, the deliverable of the CBOT 30 Year U.S. Treasury Bonds Futures are “U.S. Treasury bonds that, if callable, are not callable for at least 15 years from the first day of the delivery month or, if not callable, have a maturity of at least 15 years from the first day of the delivery month”.<sup>7</sup> Trading on the futures exchanges takes place through the open-outcry method. A large group of traders gather on the floor of the exchange (they are therefore referred to as *floor traders*) and shout out orders. Off-exchange customers need to transact through these floor traders. Starting 1998, electronic trading is possible besides the open-outcry system. Similar to the spot market, this rapidly picked up significance: over 2006 electronic trades generate about 70% of daily CBOT volume.<sup>8</sup>

There is a link between the spot and futures market for treasuries. As for any asset, arbitrage keeps the spot and futures price similar (if not identical).<sup>9</sup> An interesting finding in Fleming and Sarkar (1999) is that trading volume does not always concentrate on the same market (i.e. spot or futures). For long maturity contracts volume concentrates on the futures market. For example, for 30 year treasuries the futures market generates almost 95% of volume. For shorter maturity contracts volume concentrates on the spot market. For example, for 5 year treasuries about 75% takes place on the spot market.<sup>10</sup>

## 1.2 Datasets Analyzed and Relation Between Them

In this dissertation I look at various aspects of the market for U.S. government bonds. To this end I employ various datasets. In the first part of the dissertation, Chapter 2 and Chapter 3, I look at a high-frequency dataset of one maturity to study treasury trading. In the second part, Chapter 4, I look at a monthly dataset with a long time series of a cross-section of maturities.

---

<sup>6</sup>CME and CBOT merged on July 12, 2007. As most of the data used in this dissertation are from the pre-merger period I consider them to be separate entities and refer to them as such.

<sup>7</sup>Source: [www.cbot.com](http://www.cbot.com).

<sup>8</sup>Source: [www.cbot.com](http://www.cbot.com), “CBOT sets annual records for total volume and average daily volume”.

<sup>9</sup>For equity market evidence, see, e.g., Kawaller, Koch, and Koch (1987), MacKinlay and Ramaswamy (1988), Stoll and Whaley (1990) and Chan (1992).

<sup>10</sup>These calculations are from Fleming and Sarkar (1999), who use 1993 spot market data for ‘on-the-run’ securities (i.e. the most recently issued security in a maturity) and 1993 futures market data for the most nearby contract (i.e. the contract with the closest expiration date).

The high-frequency dataset I employ to analyze treasury trading contains all trades taking place on the CBOT trading pit of the 30Y U.S. treasury futures over the period 1994-1997. As described above, on this exchange trading takes place through floor traders. If an outside customer wants to trade he has to contact one of the floor traders to trade on his behalf. In the studies in the first part of the dissertation I benefit from the unique feature that due to regulation these floor traders have to indicate whether each trade was for their own account or on behalf of customers. This is a major advantage over spot market datasets of the U.S. treasury markets, which only contain interdealer trades. I focus on the most nearby of the outstanding futures contracts. At each point in time trading takes place in futures contracts with several expiry dates. Typically trading volume however concentrates on one of these, the nearby contract (see Fleming and Sarkar (1999)).<sup>11</sup> An advantage of looking at the nearby contract is that it is a close substitute for the spot market instrument, as Ederington and Lee (1993) point out.

The monthly dataset I use in the second part of the dissertation is constructed based on end-of-month price quotes for all treasury issues over the period 1972-2000.<sup>12</sup> There are several methods for summarizing the information in these yields into a ‘clean’ fixed maturity monthly yield dataset. Bliss (1997) provides a nice overview of the various methods. I follow his suggestion and use the unsmoothed Fama-Bliss methodology from Fama and Bliss (1987) to obtain forward rates, which is based on Center for Research in Security Prices (CRSP) government bond data. Until October 15, 1996, this database obtains its quotes on the Federal Reserve Bank of New York’s daily composite closing quotations. After this date voice-based spot market data from GovPx are used.<sup>13</sup> Fixed maturity forward rates are constructed by an iterative method, in which the forward rate for a maturity is calculated based on the forward rates for shorter maturities. These forward rates are then converted into unsmoothed Fama-Bliss zero coupon yields. The resulting dataset has as advantage that it is the same as used in the papers looking at issues related to what I do (see for example Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006)), allowing for clear comparison.

The monthly dataset relates to the above described high-frequency dataset. Besides in frequency the datasets differ in two more respects. First, the monthly dataset looks at the spot market, while the intraday dataset looks at the futures market. Second, the monthly data is constructed based on a number of outstanding treasuries, while the intraday data is only for one instrument. However, as pointed out in the previous section, there is a tight

---

<sup>11</sup>This is similar to a well-known phenomenon on the spot market: liquidity concentrates in the most recently issued security of a given maturity (referred to as *on-the-run* effect).

<sup>12</sup>I thank Francis X. Diebold for making the dataset available on his website: [www.ssc.upenn.edu/fdiebold/](http://www.ssc.upenn.edu/fdiebold/). The dataset is constructed by Professor Robert Bliss.

<sup>13</sup>Source: “CRSP Monthly Treasury U.S. Database Guide”.

link between the spot and futures market, such that there is a strong relation between the two datasets used.

## 1.3 Trading U.S. Government Bonds

In the first part of this dissertation I look at trading in U.S. government bonds. Studies that look at trading of assets and the associated costs and price behavior fit into the subfield of financial economics commonly described as *Market Microstructure*. In this section I discuss the main issues and findings in this literature, and outline the two chapters in the first part of this dissertation.

### 1.3.1 Market Microstructure

The term market microstructure comes from Garman (1976), who treats the moment-to-moment aggregate exchange behavior (the ‘microstructure’) of markets. Since this paper the term is used to describe the part of the financial economics literature that deals with price formation and the trading process. This literature has grown at a fast pace, particularly since the availability of fast computers and rich datasets. O’Hara (1995) summarizes the theoretical literature, Hasbrouck (2007) discusses the empirics, while Harris (2003) focuses more on the institutional settings. Some recent complete overview articles include Madhavan (2000, 2006), Stoll (2001) and Biais, Glosten, and Spatt (2005). This subsection is largely based on these last two papers.

At exchanges buyers and sellers do not arrive at the same time. Therefore markets rely on a subset of traders that stand ready to accommodate this asynchronous arrival. These traders are commonly referred to as *market makers*, as they can be seen to ‘make’ the market for a given security.<sup>14</sup> Arriving buyers are able to buy from market makers at the *ask-price*, sellers are able to sell to these traders at the *bid-price*. The traders arriving at the market to trade are often referred to as demanding ‘liquidity’, while the traders standing ready to trade are referred to as the ‘liquidity’ providers.<sup>15</sup> Liquidity here is interpreted as the ease with which transacting is possible. If a trader demands liquidity he requires services to transact, if a trader supplies liquidity he offers services to transact.

The difference between the bid- and ask-price, referred to as the *bid-ask spread* (or simply *spread*), is an indicator of the cost of trading.<sup>16</sup> Under the assumption of competitive

---

<sup>14</sup>Recently the trend is toward electronic markets, where traders can also submit limit orders. However, also here certain agents can be interpreted to make a market; see the discussion in the introduction of Chapter 3.

<sup>15</sup>Demsetz (1968) uses the term ‘immediacy’ to emphasize the costs incurred when requiring immediate transacting.

<sup>16</sup>This may seem as oversimplifying the many aspects of trading. Liquidity is often seen to consist



liquidity supply three factors determine the size of the bid-ask spread: (i) order processing costs (Roll (1984)), (ii) inventory risk (Stoll (1978)) and (iii) asymmetric information (Kyle (1985), Glosten and Milgrom (1985)).<sup>17</sup> Order processing (or order handling) costs are the costs liquidity suppliers incur besides the other two mentioned components of the spread, and include costs such as the costs of labor and capital needed to provide quote information, clearing and execution. Inventory costs represent the risk that the liquidity supplier is exposed to by holding a suboptimal inventory position. Asymmetric information costs come from the probability that the counterparty of a trade can be superiorly informed about the direction of the price, and that you lose money because you are on the other side of the trade.

It may seem that the costs associated to these issues are limited, perhaps not sizable enough to warrant an entire strand of academic literature or even this dissertation. It is indeed true that, looking at the treasury market as an example, the average bid-ask spread is small at about \$6 per trade. However, given the enormous volume on financial markets these small costs translate into significant amounts. Turning again to the treasury market, with more than 20,000 average daily trades with an average size of about 13 contracts, this turns into an amount of \$1,560,000 per day, or over \$7.8 million per week and over \$390 million per year! This number is for U.S. treasury trading in one maturity; the aggregate costs of all financial markets combined gives a figure in the billions!<sup>18</sup>

In the two chapters in the first part of this dissertation, Chapter 2 and Chapter 3, I look at market microstructure issues for U.S. treasuries. In Chapter 2 I focus on the third explanation of the spread: information asymmetry. I study trading in U.S. treasuries after macroeconomic news arrives at the market. I argue that differential access to customer flow is a reason for the observed increase in information asymmetry after the news releases. In Chapter 3 I study the second explanation of the spread, inventory risk, and look at the traders on the pit of the 30Y U.S. treasury futures. The large cross-section of traders with a market making role serves as an ideal sample to study market making behavior in general.

---

of three components (see, e.g., Hasbrouck (2007)): depth (is there enough quantity available near the best offered bid and ask), breadth (are there many market participants) and resilience (how long does it take for the price effects of a trade to reverse). However, as Stoll (2001) points out, these measures together converge to the bid-ask spread as the latter can be interpreted as the amount a trader needs to pay another agent to take a position and unwind it optimally.

<sup>17</sup>In the competitive case strategic liquidity suppliers can earn rents, see Biais, Glosten, and Spatt (2005, Section 2) for a recent survey of this literature.

<sup>18</sup>The spread and activity numbers in this paragraph are taken from Chapters 2 and 3.

### 1.3.2 The Importance of Customer Flow

In Chapter 2 I look at a special subset of trading days. I focus on days at which it is known in advance that information will come to the market. These days are the days at which there are the scheduled releases of macroeconomic news. For example, every first Friday of the month at 8:30 a.m. Eastern time the United States Bureau of Labor Statistics releases the employment record.

These releases of news change the equilibrium riskfree rate. Many articles report that these macroeconomic announcements are responsible for most of the observed volatility patterns in a day.<sup>19</sup> Fleming and Remolona (1999a) document a two-stage adjustment process around the news releases. In the quick first stage there is a nearly instantaneous price jump of the riskfree rate responding to the news. In the prolonged second stage volatility and trading volume remain high. Only after a longer time, typically 15 minutes after the news release but sometimes only early in the afternoon, these variables return to levels comparable to that on days without these news releases.

Green (2004) finds higher informativeness in the flow of orders after macroeconomic announcements. He shows that there is increased information asymmetry after the news releases and that in these periods order flow significantly affects the riskfree rate. This result is consistent with Brandt and Kavajecz (2004), who document informativeness of order flow on days without macroeconomic announcements, and Pasquariello and Vega (2007). It is however not obvious how to interpret information in order flow for government bonds. The classic equity market literature, such as Glosten and Milgrom (1985) and Kyle (1985), interprets this information as some traders having a private signal on a stock's future dividends.

Recent theoretical and empirical work indicates that private information contained in order flow can be broader than this classic equity market interpretation. Hasbrouck and Seppi (2001) find a common component in equity market order flow that correlates with market returns. Edelen and Warner (2001) provide further evidence of this, and document that mutual fund flows is one such factor. These studies suggest that there is information in the flow beyond an equity idiosyncratic effect. Also in nonequity markets order flow correlates with price changes. Besides the treasury market evidence discussed in the previous paragraph this is also the case on the foreign exchange market, see Evans and Lyons (2002, 2008). Recent theory suggests order flow conveys private information at the micro level, about agents' endowments and risk preferences (see Gallmeyer, Hollifield, and Seppi (2005) and Saar (2007)).

---

<sup>19</sup>See Ederington and Lee (1993, 1995), Balduzzi, Elton, and Green (2001) and Andersen, Bollerslev, Diebold, and Vega (2003, 2007). Fleming and Remolona (1997) confirm the relation between Treasury prices and public news, and conclude that the largest price shocks in the bond market over the period August 1993 until August 1994 are all caused by macroeconomic news announcements.

In Chapter 2 I focus on the part of the order flow that comes from the aggregation of private (micro-)information. The dataset I use allows to single out the part of the order flow coming from customers. This is a major distinction from the existing studies, who are only able to study interdealer flow. Interdealer flow may be related to the aggregation of micro-information, but can be different due to two reasons. First, dealers may endogenously trade based on the information in customer flow. As I show in the appendix of that chapter, this may distort the informativeness of order flow. Second, dealers may initiate trades based on a private signal they have themselves. Anand and Subrahmanyam (2007) find that the contribution of intermediaries to price discovery is unrelated to access to customer flow.

I show that customer flow is indeed informative for discovering the equilibrium riskfree rate, and that dealers speculate based on this information. Customer flow is significantly more informative in the 15 minutes interval after macroeconomic announcements, both from a statistical and economical viewpoint. I find that 25% of the explained variation in the riskfree rate after the announcement is due to customer flow. In addition I show that dealers who can observe part of the order flow speculate on this information. I find higher trading profits for *dual traders*, floor traders that both trade for own account and on behalf of customers, compared to *local traders*, floor traders that only trade for own account. This result also holds in the cross-section of dual traders: trading profits are positively and significantly related to measures of access to customer flow.

### 1.3.3 Market Makers

In Chapter 3 I turn to the market making role of the floor traders on the Chicago trading pit. Market makers are most generally defined as the traders that stand ready on an exchange to accommodate the asynchronous arrival of buyers and sellers. They do so by trading for own account: buying when sellers arrive and hoping to offset their suboptimal inventory position later against buyers (and vice versa). In this process they earn the bid-ask spread for carrying the risk of this position.

So far empirical support for models describing market making inventory behavior is weak. Studies look at inventory management by measuring mean-reversion in inventory positions. The half-life of inventory is often found to be high. For example, Hasbrouck and Sofianos (1993) report half-lives of up to two months for the New York Stock Exchange's market maker, the Specialist.<sup>20</sup> Manaster and Mann (1996) and Bjønnes and Rime (2005)

---

<sup>20</sup>Panayides (2007) argues that this is partially due to other obligations of the market maker. In addition to providing liquidity by trading for own account the Specialist has to ensure smooth price discovery through the Price Continuity rule. In times when the constraints of this rule are binding the Specialist loses, while in times when this is not the case he makes positive profits and has inventory mean-reversion with half-lives closer to what one expects.

report inventory mean-reversion, but do not find the predicted price effects of inventory.

A possible explanation for the weak empirical support for market making inventory models is, besides possible institutional restrictions, speculative activity by market makers. Madhavan and Smidt (1993) develop a model in which the market makers' trades at each time reflect two components. The first component represents the market maker managing his inventory toward a long-run desired inventory level. The second component relates to short-run speculative trades. Ignoring this second part results in drawing incorrect inferences from inventory time series.

Using the large cross-section of over 3,000 traders on the Chicago trading pit my results provide direct evidence of speculative position taking by market makers. I find empirical support for both components of the Madhavan and Smidt (1993) model. Inventory reverts to a long-run position, as floor traders close out the day with a zero inventory position. In addition I find that market makers initiate trades that increase their inventory position. Relating the extent of this active position taking to profits from trading there is a positive and significant correlation.

The results add to the debate whether the market maker is indeed a mere uninformed liquidity supplier or an informed speculator. Manaster and Mann (1996) and Bjønnes and Rime (2005) interpret their findings as being consistent with a speculative role of market makers. Anand and Subrahmanyam (2007) show that intermediaries, a subset of the market makers that I consider, can be informed. Chakravary and Li (2003) however find no evidence in favor of speculation. I am the first to provide direct evidence of profitable market maker speculation.

In the analysis I come across the challenge of determining whether the buying or selling party initiated a trade (the so-called *signing* of trades) when no quotes are available. For financial markets where quotes are available algorithms to identify whether the buying or selling party initiated a trade are available, see, e.g., Lee and Ready (1991). For markets without quotes this task becomes (even) more challenging, as besides the ignorance of initiating party also the midquote is challenging to estimate. Hasbrouck (2004) suggests a Bayesian methodology to perform this task, based on the Roll (1984) model. However, as it is a simulation based methodology it is numerically intensive. With over 40 million observations this method takes too long for my purposes.

I suggest an alternative implementation of the Roll (1984) model, which relies on time series methods for models in *state space form*. In state space models the observed time series are explained by latent factors. Assuming a normally distributed measurement error and a Gaussian transition process (such as a Vector Autoregression process) for these unobserved factors a filtering algorithm known as the *Kalman filter* can be employed.<sup>21</sup>

---

<sup>21</sup>For more on state space form models see, e.g., Harvey (1989) and Durbin and Koopman (2001).

By its nature the Roll (1984) model fits well into the state space framework: the observed price is explained by an unobserved true (or ‘efficient’) price and a time series of discrete trade initiation variables. It is however outside the scope of the linear Gaussian state space model. To overcome this difficulty I follow techniques macroeconomists use to model time series which process depends on whether the economy is in a state of recession or expansion. These techniques, detailed in, e.g., Kim and Nelson (1999), exploit the property that conditional on observing the discrete variable the model is linear and Gaussian, and complete the filter by making inferences about the probability terms for the state of the economy.

## 1.4 Yields of U.S. Government Bonds

The second part of the dissertation consists of one chapter, Chapter 4. I turn to the *term structure*, or *yield curve*, of government bonds: the cross-section of yields for bonds with different maturities over time. In this section I first provide an overview of the developments in the term structure literature and then outline the fourth chapter in this dissertation, in which I improve a popular model for the term structure.

### 1.4.1 Modeling the Yield Curve

The unique structure of the yield curve has interested researchers since long. Bond yields for different maturities are related: yields on bonds with a long maturity can be seen as risk-adjusted expectations of future yields on bonds with a short maturity. Due to this relation it is possible to summarize the information in the yield curve with a few common factors driving all these rates. Still a lot is unknown about the exact relation between the yields for different maturities and what is exactly driving the bond market. The model to beat for both fitting and forecasting the yield curve still seems to be the random walk: the best prediction of next month’s yield curve is the current yield curve.

There is great significance in understanding what moves the bond market. Piazzesi (2003) mentions four reasons why researchers should look at the yield curve. First, due to the relation between yields on bonds with a long and short maturity the yield curve contains information about the future path of the economy. Decisions of firms, consumers and policy makers all depend to some extent on forecasts based on the yield curve. Second, the working of the yield curve matters for monetary policy. In most countries the central bank can most effectively influence short term rates. For certain decisions the long term rate seems to be more informative however, such as for consumers buying a house. The relation between these yields provides insights into how movements at different maturities relate to each other. Third, to decide on optimal (government) debt policy it is necessary

to find out how the entire yield curve responds to an increase in supply of bonds for a given maturity. Fourth, the yield curve matters for derivative pricing and hedging. The price of many derivatives depends on the entire yield curve (for example futures and interest rate options). Banks need to hedge their interest rate risk exposure due to short-term interest payments and long-term interest collections.

The earliest interest rate models are all economic in nature. Vasicek (1977) and Cox, Ingersoll, and Ross (1985) pioneered the class of models commonly referred to as affine term structure models. In here an affine function of an unobserved factor, the short rate, explains the yield curve. Duffie and Kan (1996) generalize this literature; Dai and Singleton (2002) characterize the set of admissible and identifiable models. Another significant development came in the 1990's, when the class of no-arbitrage models was introduced. These models focus on fitting the term structure at a given point in time to ensure no arbitrage opportunities exist, see Hull and White (1990) and Heath, Jarrow, and Morton (1990).

The model I look at is the celebrated Nelson and Siegel (1987, NS) model for the yield curve. This method decomposes the entire yield curve at a point in time into three unobserved components. These components are interpreted as the level, slope and curvature of the yield curve. The method is popular amongst both practitioners and central banks. Of the 13 central banks studied in the Bank for International Settlements (BIS) report "Zero-coupon yield curves: technical documentation", 9 use the Nelson and Siegel (1987) model to estimate zero-coupon yield curves. Whilst being statistical in nature it has the advantage that the components carry a clear interpretation. Two recent publications by Francis X. Diebold have strengthened the importance of the NS model. Diebold and Li (2006) show that forecasts obtained from the NS model outperform competing economical and statistical models. Diebold, Rudebusch, and Aruoba (2006) argue that the NS model is in fact a model in state space form, as it explains observed time series with latent factors. I follow this approach of analyzing the Nelson and Siegel (1987) model, which is commonly referred to as the dynamic Nelson Siegel model.

## 1.4.2 Extensions to the Nelson-Siegel (1987) Model

Parameter estimation in the dynamic Nelson and Siegel (1987) model relies on two simplifying assumptions, that I relax in Chapter 4. First, researchers assume that the parameter that determines the loadings of the latent factors onto the maturities is constant. I propose a model in which the factor loadings parameter is made time-varying. Initially I investigate whether there indeed is evidence that the factor loadings parameter varies over time by modeling it with a spline. As this analysis suggests this is indeed the case, I model it in an even more flexible way, by adding the factor loadings parameter to the

state. As this model specification is nonlinear, standard Kalman filter recursions can not be used. Following the extended Kalman filter literature, see, e.g., Anderson and Moore (1979), I use an approximation of the nonlinear model that uses Taylor series. The second assumption in the dynamic Nelson-Siegel model I relax is volatility being constant over time. Similar to the factor loadings I initially study this using a cubic spline, which provides significant evidence of time-varying volatility. I then model this using the well-known class of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, as pioneered by Engle (1982) and Bollerslev (1986).

I show that relaxation of these two assumptions contributes significantly to the model both from a statistical and economical point of view. Using likelihood based tests I find significant improvements in model fit for both making the factor loadings parameter and volatility time-varying. Introducing time-varying volatility improves the accuracy of forecast confidence intervals. The time-varying factor loadings term structure model provides important economic insights into new developments in the literature. Christensen, Diebold, and Rudebusch (2007) develop an arbitrage-free Nelson-Siegel model, defining a new class of term structure models. A condition for the dynamics to be arbitrage-free is constancy of the factor loadings parameter. Diebold, Li, and Yue (2007) develop a global yield curve model, constructed based on the assumption of constant factor loadings. The results in Chapter 4 provide evidence against this.

# Chapter 2

## Customer Flow, Intermediaries, and the Discovery of the Equilibrium Riskfree Rate

This chapter is based on Menkveld, Sarkar, and Van der Wel (2007).

### 2.1 Introduction

In frictionless markets, asset prices reflect public news instantaneously. We should therefore observe price changes only on announcements. Empirically, we see asset prices also change in the absence of announcements. This observation motivates the introduction of several market frictions to improve our understanding of asset price behavior and a prominent one is asymmetric information. That is, information is distributed asymmetrically across agents in the economy and the equilibrium price is learned through iterating over price quotes and updating based on the (aggregate) order imbalances that these prices provoke.<sup>1</sup> The market aggregates private information.

Recent evidence indicates that such private information is broader than the classic equity market interpretation of a private signal on a stock's future dividends. In the equity market itself, for example, the order flow is informative beyond the conjectured idiosyncratic effect as it contains a common factor that correlates with daily market returns.<sup>2</sup> Furthermore, order flow in nonequity markets, such as the currency and the

---

<sup>1</sup>In *Eléments d'économie politique pure*, Walras (1889) first introduces the idea of tâtonnement, where agents submit buy (sell) orders when prices are low (high). Prices adjust to reflect the order imbalance until there are no additional orders. The equilibrium value has been discovered.

<sup>2</sup>See Hasbrouck and Seppi (2001). Edelen and Warner (2001) find correlated mutual fund flow to be part of this factor.



treasury market, correlates significantly with permanent price changes.<sup>3</sup> Inspired by asset pricing models, the literature proposes that order flow conveys private information about the agent's optimization problem at the micro level, including her preferences and her endowments.<sup>4</sup>

Our main goal is to identify this source of information in the order flow—the part that captures the aggregation of private (micro) information from the economy. The existing studies cannot identify this source for two reasons. First, they are based on interdealer flow which reflects customer order flow,<sup>5</sup> but also contains dealer-initiated trades. These trades might reflect information other than the type we aim to identify, as some dealers might, for example, be superior information processors.<sup>6</sup> Also, dealers might initiate trades based on privately observing their customer identity, which endogenously biases dealer flow informativeness as a measure of customer flow informativeness. We elaborate on this flow-based speculation argument below. Second, existing studies do not control for a potential reverse causality due to “feedback trading” on stale prices. That is, public information arrives that causes the efficient price to change and, at the same time, agents trade against outstanding (stale) quotes that stand in the way of price adjustment. Hence, in the interval we witness a price change and trades in the direction of the change.

We turn to trading in 30-year treasury futures following macroeconomic announcements as an appropriate laboratory to identify private information aggregation in the economy for two main reasons. First, intermediaries have to report for-customer trades, which allows us to remove intermediary-initiated flow from the net order flow. Second, we observe the public signal—the macro “surprise”—and can therefore identify and remove the part of order flow that is feedback trading.

Our results show that off-market customer flow is important for discovering the equilibrium riskfree rate. Relative to nonannouncement days, we find that customer flow is significantly more informative on price changes in the first 15 minutes after an announcement. Economically, the contribution of customer flow to price discovery is substantial as, after removal of the “feedback trading” part, it accounts for one-fourth of the (explained)

---

<sup>3</sup>See, e.g., Evans and Lyons (2002, 2008) for the FX market and Brandt and Kavajecz (2004), Green (2004), and Pasquariello and Vega (2007) for the treasury bond market.

<sup>4</sup>See, e.g., Evans and Lyons (2002), Gallmeyer, Hollifield, and Seppi (2005), and Saar (2007).

<sup>5</sup>Dealers typically accommodate a customer order completely and then actively unload their inventory in the interdealer market. For example, a customer sell order therefore creates a series of sell orders in the aggregate interdealer market. See Lyons (1997) for a discussion of such “hot potato trading”.

<sup>6</sup>Anand and Subrahmanyam (2007) find that for the Toronto stock exchange the most informative trades are initiated by intermediaries unrelated to how much access they have to customer flow. Pan and Poteshman (2006) and Kurov and Lasser (2004) show that one potential source of such information is proximity to the aggregate order flow. By nature, this information is a short-lived (as dealers typically go home flat) whereas our focus is on the “long-lived” macro information that markets produce when aggregating micro customer flow.

riskfree rate change.<sup>7</sup> We further find that this informativeness increase is significantly larger for those months in which the dispersion in analyst forecasts is high. This suggests that intermediaries rely more heavily on customer flow at times of high disagreement on macro fundamentals.

Our finding provides further insight into the price discovery process in treasury markets. Prior studies also address the issue, but, unlike us, they use GovPx data that only covers interdealer flow and they are therefore unable to identify the origination of the information. For example, Ederington and Lee (1993) and Fleming and Remolona (1999a) report a strong instantaneous response of treasury bond prices to an announcement, but also increased volatility in the minutes after the announcement. Green (2004) documents that in the first 15 minutes after the announcement treasury returns show increased sensitivity to order flow relative to the same time interval on nonannouncement days. Pasquariello and Vega (2007) find that the correlation increases with the dispersion in analyst forecasts.

In addition to customer trade identification, our treasury futures sample has some attractive features. First, we do not need an algorithm to sign trades, as the data identifies for all customer transactions whether these customers buy or sell. Second, it is comprehensive as 30Y treasury futures capture 95% of the trading volume in the spot and futures markets for this maturity (see Fleming and Sarkar (1999)).

The second part of the chapter generates further support for customer flow informativeness through an analysis of own-account profitability in the cross-section of intermediaries. The key idea is that the intermediary benefits from *privately* observing the identify of her customer. The futures exchange forbids any activity for own-account *ahead* of executing a client order on the floor, i.e. broker-dealers cannot frontrun a client order nor execute it (partially) against their own account (see Grossman (1989, p.6)).<sup>8</sup> As the origination of the order is not revealed in the trading process, after executing her client order the intermediary still benefits from having privately observed her customer identity by trading for own-account (on the floor) in same the direction as her customer if her customer was informed (“piggyback”) and in the opposite direction if her customer was uninformed. The aggregate (i.e. customer plus own-account) net trade in the interdealer flow therefore amplifies the information part and reduces the noise part of the customer order. Zero-profit market makers rationally charge higher price impacts to protect themselves against such flow-based speculation, which thus endogenously biases interdealer

---

<sup>7</sup>Specifically, we decompose the *increase* (relative to nonannouncement days) of the (explained) riskfree rate variance in the 15 minutes after the announcement and find that 76.0% is instantaneous and 24.0% is learned from customer flow.

<sup>8</sup>This institutional feature makes an alternative explanation for own-account profitability based on bargaining power in the spirit of Green, Hollifield, and Schürhoff (2007) unlikely.

flow informativeness as a measure of customer flow informativeness.<sup>9</sup>

We exploit the large cross-section of 3,382 intermediaries and relate own-account profitability to customer flow access to provide direct evidence on flow-based speculation. We find two key results. First, we report that own-account profitability is higher for intermediaries who also trade for customers (“duals”) relative to those who do not (“locals”). The benchmarking against locals serves to control for the increased cost of market-making in the volatile postannouncement period.<sup>10</sup> Second, we exploit the cross-section of duals to show that their own-account profitability increases with access to customer flow, where we control for volatility, competition, and the macro “surprise”. Intermediaries therefore appear to trade profitably on the information in customer flow, which feeds our earlier concern that (part of) the increased sensitivity of riskfree rate change to (aggregate) interdealer flow might be endogenously generated.

We entertain the alternative explanation that intermediaries with superior trading skill are likely to attract more customers (see, e.g., Grossman (1989)), which makes the correlation between own-account profitability and access to customer flow entirely spurious. To control for skill, we compare an intermediary’s own-account profitability on announcement days where she has access to customer flow relative to announcement days where she does not and find significantly increased profitability on days where she has access to customer flow. Furthermore, we find that on the announcement days that she does not trade for customers, her own-account profitability is not significantly different from own-account profitability of locals. These results rule out that exceptional trading skill drives a dual trader’s increased profitability.<sup>11</sup>

We further analyze postannouncement trading to firmly establish that the increased sensitivity of riskfree rate changes to customer flow reflects information. We realize that in inactive markets any regression of price change on signed flow might pick up a transitory price effect to compensate for the cost of market-making. For example, the increased sensitivity might reflect that risk-averse dealers require higher compensation for carrying inventory through time on increased postannouncement volatility. We consider this non-

---

<sup>9</sup>The idea that intermediaries benefit from discriminating informed from uninformed flow is well-established in the literature. Market makers cream-skim uninformed flow (see, e.g., Benveniste, Marcus, and Wilhelm (1992), Easley, Kiefer, and O’Hara (1996) and Chung, Chuwonganant, and McCormick (2004)). Brokers trade along informed flow (Fishman and Longstaff (1992)) or against uninformed flow (Roell (1990), Madrigal (1996)). Appendix 2A illustrates the idea in a Kyle (1985) setup and includes a rational response of the informed customer who reduces her order size in anticipation of the intermediary’s speculation.

<sup>10</sup>We find supportive evidence for such increased cost of market-making as a local’s (gross) own-account profitability per round-trip trade is higher in the first 15 minutes after an announcement relative to profitability in the same period on nonannouncement days. We also find a significantly increased bid-ask spread in this period.

<sup>11</sup>We interpret trading skill broadly to include an ability to quickly process and interpret macro news as in Kim and Verrecchia (1994, 1997).

information explanation unlikely for our five-minute regressions in what is a very active market. That is, for an average announcement day five-minute interval, 172.9 intermediaries generate 595.9 transactions. Moreover, if we regress interest rate changes on only those customer orders that trade through intermediaries who do not trade for own-account that day, we find *unchanged* sensitivity.<sup>12</sup> Consistent with the flow-based speculation, it seems that intermediaries endogenously choose to trade for own-account on recognizing informed customers in their total customer flow.

Finally, we contribute to the dual-trading literature. Chakravary and Li (2003) study eight CME futures contracts and find that dual traders supply liquidity and actively manage inventory. Manaster and Mann (1996) corroborate these findings in their CME futures study, but, much to their surprise, also report a positive correlation between signed inventory and the intermediary reservation price. They conclude that intermediaries are not “passive order-fillers, . . . but active profit-seeking individuals with heterogeneous levels of information and/or trading skill”. We establish that one channel is access to informative customer flow. Most related to our study is Fishman and Longstaff (1992) who propose a model to illustrate that the decision to trade for own-account is endogenous, i.e. the intermediary does so if she has private knowledge on the composition of her customer order flow. Our study differs in three ways. First, we focus on trading in the wake of a macro-announcement so as to control for a reverse causality caused by “feedback trading”. We establish that customer flow is indeed informative. Second, we exploit a large cross-section of duals to establish that access to customer flow is a key determinant of own-account profitability. Third, we study a much larger sample (42.5 million trades in 4 years vs. 305,982 trades in 15 days) and benefit from statistical power which, for example, allows us to reject the alternative explanation based on trading skill, which could not be rejected in Fishman and Longstaff (1992).

The remainder of the chapter is organized as follows. Section 2.2 discusses the institutional background, the data, and provides summary statistics. Section 2.3 studies customer order flow informativeness on announcement days (relative to nonannouncement days). Section 2.4 calculates the intermediary’s own-account trading profit and relates it to access to customer flow. Section 2.5 analyzes who effectively pays the intermediary’s increased profitability. Section 2.6 concludes.

---

<sup>12</sup>This is not an order size effect, as customer orders in this subset are larger than the average customer order.

## 2.2 Background, Data, and Summary Statistics

### 2.2.1 Background

We analyze four years (1994-1997) of trading in 30Y treasury futures at the Chicago Board of Trade (CBOT). At the time, this contract is one of the most liquid securities with 485.2 trades every five minutes on nonannouncement days and even more on announcement days.<sup>13</sup> Almost all trading is floor trading from 8.20 a.m. to 3.00 p.m. Eastern Time (ET), although after-hours electronic trading volume had been growing. Trading occurs in a pit by means of the so-called open outcry method. Floor traders negotiate prices by shouting out orders to other floor traders, indicating quantity and trade direction through hand signals. Other floor traders bid on the orders, also using hand signals. Once filled, an order is recorded separately by both parties to a trade. At the end of the day, the clearinghouse settles trades and ensures that there is no discrepancy in the matched trade information.

After a criminal inquiry in 1989, the Commodity Futures Trading Commission (CFTC) –the main regulatory body of futures exchanges– continues to allow dual trading, but tightens surveillance. An FBI sting operation at the CBOT and the Chicago Mercantile Exchange (CME) finds that brokers (including dual traders) are cheating customers and leads to dozens of arrests. In 1992, Congress mandates that futures markets keep audit trails. The CFTC pressures both CBOT and CME to supply the information with the threat of a dual trading ban, in case the exchanges fail to comply.<sup>14</sup> Today, dual trading continues to be allowed in most futures markets. The exceptions are some CME futures contracts, mostly those with a history of high volume.

### 2.2.2 Data

**Futures data.** We benefit from the CFTC audit trail data to discriminate customer trades and own-account trades in the 30Y treasury futures market. Each transaction record contains: contract traded (i.e. the expiration month); time;<sup>15</sup> buy or sell indicator; number of contracts traded; price; identification number for the floor trader who executes the trade; and a customer type indicator (CTI code). These CTI codes are defined in CFTC rule 1.35(e) as: CTI1 is a trade for own account; CTI2 is a trade for clearing

---

<sup>13</sup>345.9+112.3=485.2, see Table 2.2. Note that this table double-counts out of necessity, as we also report trade activity by trader type. We double-count throughout the chapter in order to be consistent.

<sup>14</sup>See, e.g., “CFTC demands tighter controls”, *Financial Times*, 8/13/96.

<sup>15</sup>Traders report time in 15-minute brackets and an exchange algorithm, known as computerized trade reconstruction (CTR), times the trade to the nearest second. Although noisy, we believe the CTR time is fairly accurate due to Congress and CFTC pressure to provide high-quality data for surveillance. Others have used CTR time for analysis, e.g. Fishman and Longstaff (1992) and Manaster and Mann (1996).

member's house account; CTI3 is a trade for another member present at the exchange floor, or an account controlled by such other member; CTI4 is a trade for (off-exchange) customers. Consistent with earlier studies<sup>16</sup> we restrict attention to CTI1 and CTI4 trades as they represent almost all trading volume.

We focus on the nearby futures contract and apply a number of filters to prepare the data for analysis. We choose to analyze the nearby contract, as it is a very close substitute for the underlying spot instrument. Consequently, we feel that our results generalize to spot rates (see also Ederington and Lee (1993, p.1164)). We apply the following filters. We eliminate spread trades (e.g., butterfly spread trades). We remove trades that occur at unusually low prices (primarily in May 1997). We remove trades that show an unusual transaction return of more than 0.25% followed by a transaction return in the opposite direction of more than 0.25%. We expect these trades to suffer from a serious timing error. These filters eliminate 1.48% of all CTI1 and CTI4 transactions. The final sample includes 42.5 million observations.

**Macro announcements** We follow Green (2004) and use the International Money Market Services (MMS) data on expectations and realizations of the most relevant 8:30 U.S. macro announcements. Table 2.B1 provides an overview of the relevant announcements in this dataset. We are careful to remove days with macro announcements scheduled at a time later in the day (e.g. 9:15 or 10:00) to create benchmark nonannouncement days that are not contaminated by macro news trading.<sup>17</sup> We further remove (i) days when either the realized value or the expectation is missing, (ii) days when the Fed announces earlier or later relative to schedule, (iii) days with unexpected Fed announcements, (iv) days where the market is partially or completely closed.<sup>18</sup>

Table 2.1 lists the 15 macro announcements included in the sample and reports their frequencies. In total, the sample contains 377 announcement days and 350 nonannouncement days. In addition to an analysis of all announcement days, we also analyze the subgroup of most influential announcements—nonfarm payroll employment, PPI, and CPI—but also “nonfarm payroll” as a separate group as it is the single most important announcement (see also, e.g., Green (2004, Table III)).

Consistent with previous studies, we define announcement surprises as the difference between realizations and expectations (see, e.g., Green (2004) and Pasquariello and Vega (2007)). More specifically, since measurement units vary across macro variables, we standardize the surprises by dividing each of them by their sample standard deviation. The

---

<sup>16</sup>E.g., Fishman and Longstaff (1992), Manaster and Mann (1996), and Chakravary and Li (2003).

<sup>17</sup>This also removes days with e.g. both an 8:30 and a 10:00 announcement. The remaining announcement days are therefore 8:30-only announcement days.

<sup>18</sup>These days are 4/1/94, 4/5/94, 9/14/94, 8/26/96, 2/26/97, and 2/27/97.

**Table 2.1: Announcement and Nonannouncement Days**

This table shows the number of announcement and nonannouncement days in our sample, and the frequency of each announcement. The data on macroeconomic announcements is from the International Money Market Services (MMS). The announcement days are days on which there is an 8:30 announcement and no other announcement in the morning (i.e., no 9:15 and 10:00 announcements). Nonannouncement days are days on which there are no announcements at all in the morning. There are three groups of announcement days: the first group contains all 8:30 announcements, the second group consists of the important announcement types (Nonfarm Payroll Employment, PPI, and CPI), and the third group contains only the Nonfarm Payroll Employment announcements. We exclude days when either the realized value or the expectation is missing, days on which the Fed made an earlier than usual or an unexpected announcement, the day on which the Durable Goods Orders figure was announced at 09:00 or 10:00, two days on which the market closed at 11:00 (4/1/94 and 4/5/96) and four days on which the market closed for a part of the day (9/14/94, 8/26/96, 2/26/97 and 2/27/97).

Panel A: Announcement vs. Nonannouncement Days					
	1994	1995	1996	1997	Total
All Trading Days	253	250	252	250	1,005
Nonannouncement Days	84	91	88	87	350
All Announcement Days	98	90	89	100	377
Nonfarm, PPI, and CPI	27	26	25	27	105
Nonfarm Payroll Employment	9	8	7	10	34
Panel B: Announcement Types and Frequencies					
Announcement Type	1994	1995	1996	1997	Total
GDP Advance	3	4	1	4	12
GDP Preliminary	3	1	1	2	7
GDP Final	3	0	5	2	10
Nonfarm Payroll Employment	9	8	7	10	34
Retail Sales	9	11	9	12	41
Personal Income	5	3	5	4	17
Personal Consumption Expenditure	5	3	5	4	17
Durable Goods Orders	11	11	8	7	37
Business Inventories	0	0	0	7	7
Net Exports	12	10	11	11	44
Producer Price Index	11	11	11	10	43
Consumer Price Index	7	7	7	7	28
Housing Starts	11	9	10	9	39
Index of Leading Indicators	5	2	6	6	19
Initial Unemployment Claims	40	37	36	43	156

surprise  $S_{kt}$  of type  $k$  on day  $t$  is therefore

$$S_{kt} = \frac{R_{kt} - M_{kt}}{\sigma_k}, \quad (2.1)$$

where  $R_{kt}$  is the announced value,  $M_{kt}$  is its MMS median forecast that proxies for the market expectation, and  $\sigma_k$  is the sample standard deviation of  $(R_{kt} - M_{kt})$ . Equation (2.1) facilitates meaningful comparisons of how the 30Y riskfree rate responds to the different types of macro news. Operationally, we estimate these responses by regressing 30Y treasury futures price changes on the surprise  $S_{kt}$ . We note that since  $\sigma_k$  is constant for any indicator  $k$ , the standardization does not affect the statistical significance of the response estimates nor the fit of the regressions.

### 2.2.3 Summary Statistics

Figure 2.1 plots transaction prices and customer volume imbalance on a representative macro announcement day. It illustrates some trading characteristics that will turn out to be true more generally. First, the 8:30 announcement leads to an instantaneous price change of almost 1%. Second, right after the announcement we observe increased (signed) customer volume imbalances which level off after roughly 15 minutes. Third, in this time period we observe large price changes that seem to correlate with the signed customer flow. These findings are consistent with earlier papers (see, e.g., Fleming and Remolona (1999a) and Green (2004)).

**Intraday patterns.** Figure 2.2 presents the intraday patterns of volatility, the bid-ask spread, and volume. We use all 377 announcement days and 350 nonannouncement days to calculate the value for each 15-minute interval and we estimate the patterns through regressions. We use GMM for all regressions in the chapter and we use robust Newey-West standard errors (where we allow for autocorrelation up to three lags). We plot our estimates and we add a solid dot when the difference between announcement and nonannouncement days is significant at the 99% level.

Panel (A) shows that on announcement days, volatility is unchanged ahead of announcement, but significantly higher in the first half of the trading day with a clear peak in the first 15 minutes after the announcement. To avoid a bias due to the bid-ask bounce, we define volatility as the standard deviation of only customer buy transaction prices<sup>19</sup> (see also Manaster and Mann (1996)). We find a significant spike in volatility of roughly 300% in the 15 minutes after the announcement. For the rest of the day, volatility levels remain increased relative to nonannouncement days, but the increase is substantially

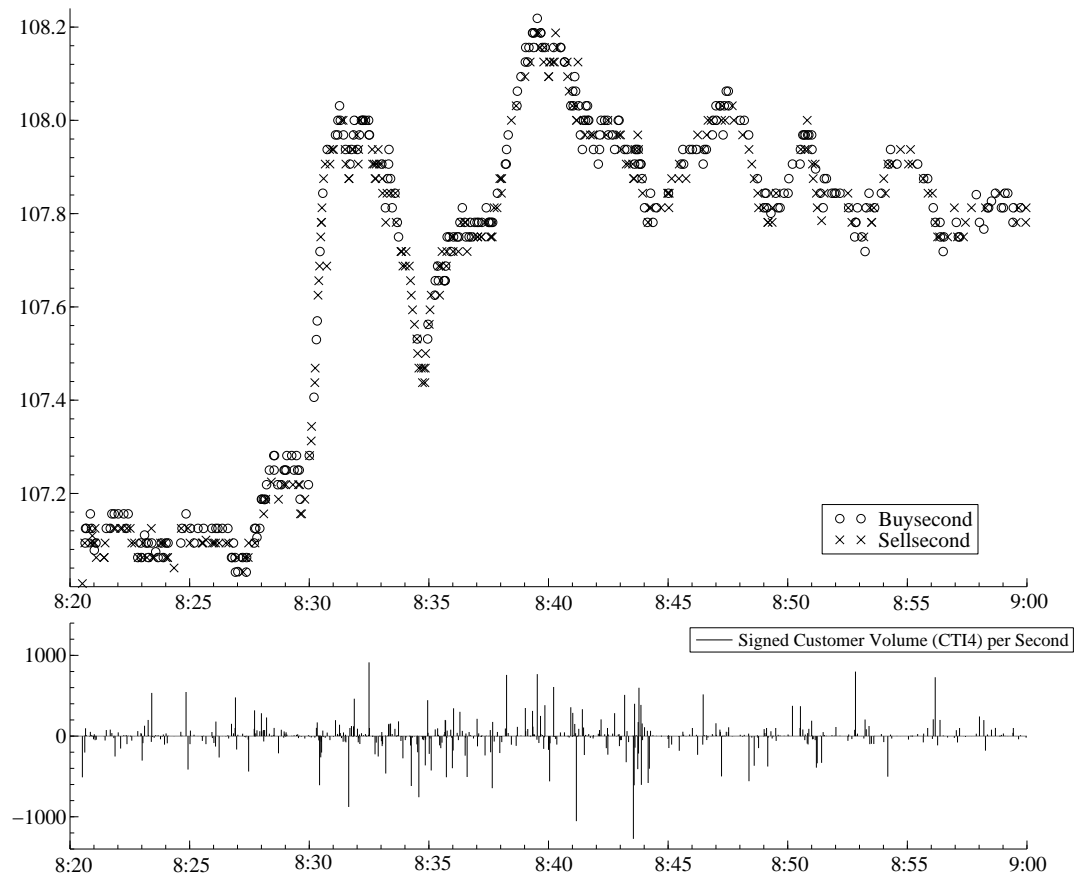
---

<sup>19</sup>Operationally, to minimize missing values, we calculate two standard deviations, one based on customer buys, the other on customer sells. We take the maximum if both are available.



### Figure 2.1: 30Y Treasury Futures Trading on an Announcement Day

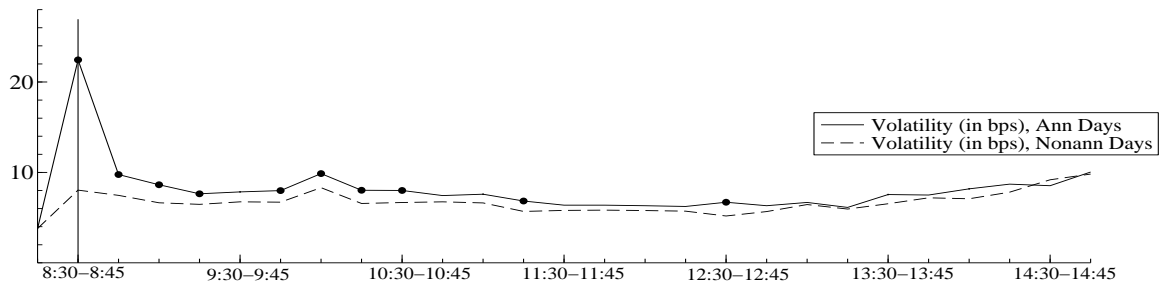
This figure depicts the prices of the 30Y treasury bond futures listed on the Chicago Board of Trade (CBOT) in the interval 8:20-9:00 on May 3, 1996. On this day there was an 8:30 Nonfarm Payroll Employment announcement. The top graph plots the volume-weighted average price for the second, where we use a circle (cross) if customer buying volume exceeds (falls below or equals) customer selling volume. The bottom figure plots the signed customer volume (CTI4) for every second.



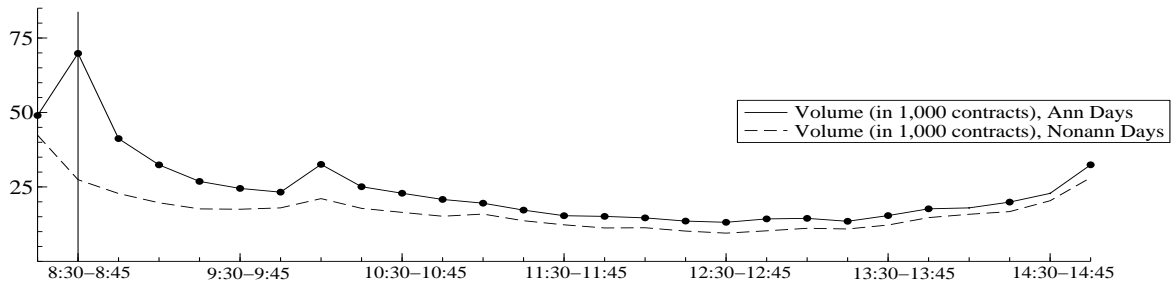
**Figure 2.2: Intraday Trading Patterns**

These figures depict intraday pattern of volatility (A), volume (B), and the bid-ask spread (C), based on fifteen minute intervals. The solid (dashed) lines show the intraday pattern for announcement (nonannouncement) days, the solid vertical lines represent the 8:30-8:45 announcement interval. A closed circle indicates a significant difference between announcement and nonannouncement days at the 99% level.

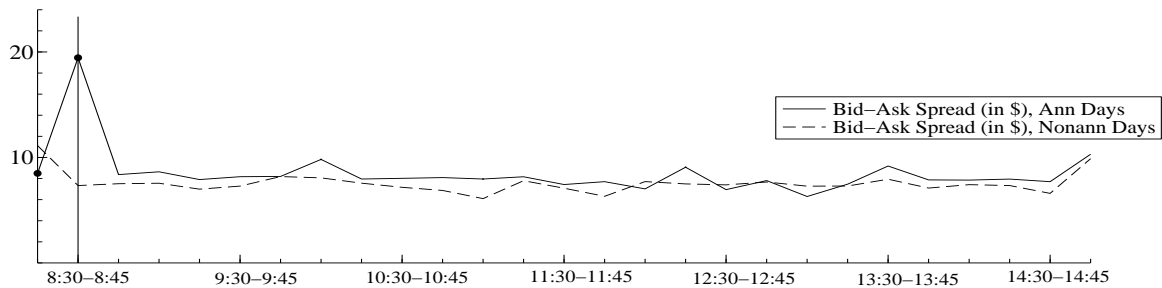
(A) Volatility (in bps)



(B) Volume (in 1,000 contracts)



(C) Bid-Ask Spread (in \$)



lower as it never exceeds 25%. The increase is statistically significant only in the early half of the day.

Panels (B) and (C) show a significant volume increase throughout the trading day on a significantly increased bid-ask spread only in the first 15 minutes after the announcement. We report aggregate volume (i.e. customer plus own-account volume) and find its increase to be similar in magnitude to the volatility increase. We estimate the bid-ask spread as the difference between the average (volume-weighted) customer buy price and the average customer sell price (see also Manaster and Mann (1996)). We only find a significant increase at the 99% level in the first 15 minutes after the announcement. Economically, the increase is substantial as it exceeds 120%. We also find significantly increased volume ahead of the announcement on a significantly lower bid-ask spread, but these effects are small economically relative to postannouncement trading.

All in all, these patterns are consistent with Green (2004) who documents increased informed trading only for the first 15 minutes after the announcement. Our volatility and bid-ask spread patterns are consistent. The increased volume in the remainder of the day might reflect inventory-sharing trades among market makers who are pushed into suboptimal positions in the first 15 minutes.

**Customer vs. own-account trades.** As it is our objective to further understand these trading patterns, we exploit our sample's unique feature that it discriminates customer trades and own-account trades. We follow the literature (see, e.g., Fishman and Longstaff (1992)) and disaggregate volume for each day according to (i) whether the intermediary trades for customers and own-account that day<sup>20</sup> and (ii) whether the trade is a customer trade or an own-account trade. We label the order flow accordingly, i.e. we get four categories:

1. Customer trades through duals, i.e. customer trades through an intermediary who also trades for own account
2. Own-account trades by duals, i.e. own-account trades of an intermediary who also trades for customers
3. Customer trades through brokers, i.e. customer trades through an intermediary who does not trade for own account
4. Own-account trades by locals, i.e. own-account trades of an intermediary who does not trade for customers

---

<sup>20</sup>We use a 2% error margin for classification (i.e., no own-account trades means less than 2% of the intermediary's trades are for own-account) as CFTC and exchange staff acknowledge the presence of error trades and consider the 2% filter reasonable (see Chang, Locke, and Mann (1994)).

We emphasize that an intermediary's label as broker, local, or dual is based on her activity on a particular day and, throughout the sample, an intermediary can therefore have broker days, local days, and dual days.

Table 2.2 presents trade statistics for announcement as well as nonannouncement days. Panel A testifies to the high activity in the 30Y treasury futures market. On nonannouncement days, we find that, on average, in a five-minute interval 42.4 intermediaries trade customer orders and 116.8 trade for own-account. They generate 112.3 and 345.9 trades, respectively. On announcement days, the number of active traders increases by approximately 20% and the number of transactions by 30%.

Panel A further disaggregates activity according to intermediary type and finds, for nonannouncement days, that the majority of active intermediaries acts as local (65%),<sup>21</sup> followed by dual (28%), and broker (7%). Clearly, dual activity continues to be substantial in the aftermath of the 1992 Congress mandate (see Section 2.2.1), in particular with regard to customer trades. For the average five-minute interval, 34.5 duals carry out an aggregate 90.9 transactions for their customers vs. 7.9 brokers who carry out 21.4 customer transactions.<sup>22</sup> Trade size is larger for brokers, but even in terms of volume duals carry out most customer orders. Furthermore, the bid-ask spread is higher for customer trades through duals vs. brokers, which is a first indication that their order flow includes the informed customer orders.

For announcement days, activity is higher across all trader types, trades are larger, and bid-ask spreads are higher. These changes appear to be proportional across trader types, so that on a relative basis the nonannouncement day characterization of trading remains true for announcement days. The same goes for the first 15 minutes after the announcement with the exception that the proportional increase in customer trades is larger than the increase in own-account trades.<sup>23</sup>

Panel B presents the mean and standard deviation of five-minute signed customer volume in the 15 minutes after a macro announcement. These statistics are useful for our main analysis in the next section where we explore customer flow as an explanatory variable for 30Y treasury returns. The panel shows that, on average, customers are net buyers after an announcement but their net flow has a very large standard deviation relative to its mean. For the category of all announcement days, for example, we find that net customer flow is 0.142 with a standard deviation of 1.282. We decompose customer flow and find that dual-intermediated net flow is larger than broker-intermediated net

---

<sup>21</sup> $100\% * 81.4 / (81.4 + 35.4 + 7.9)$ .

<sup>22</sup>Note that we find a slight difference between the number of duals active based on own-account counting (35.4 per five-minute interval) or for-customer counting (34.5). This difference is due to the counting procedure, as, apparently, a dual's own-account trading is more spread out in the day, while her customer flow concentrates in some intervals.

<sup>23</sup>These results are not included here but are available upon request.

**Table 2.2: Trade Statistics by Trader Type and Signed Customer Volume**

In Panel A, we show the average number of traders active, number of transactions, trade size (in #contracts) and bid-ask spread (in \$) per five minute interval for the 30Y treasury futures listed on the Chicago Board of Trade (CBOT) on both announcement and nonannouncement days. The averages are taken over the full day in five minute intervals, we show the variables for different trader types. We define a floor trader to be a local (broker) on a day if the proportion of volume for her own account, as a ratio of total (own-account + customer) volume, is greater than 98% (smaller than 2%). A floor trader is a dual on a day if this proportion is greater than or equal to 2% but less than or equal to 98%. In Panel B we show statistics for the signed customer volume (CTI4, *Customer Flow*, in 1,000 contracts). We show mean and standard deviation (*St Dev*) for five minute intervals as calculated for 8:30-8:45. We split the aggregate signed customer volume to the dual- and broker-intermediated parts.

Panel A: Trade Statistics by Trader Type (five min avg, full day)						
	Own Account (CTI 1)			For Customer (CTI 4)		
	Ann	Nonann	Ratio	Ann	Nonann	Ratio
#Traders Active	138.3	116.8	1.18	50.9	42.4	1.20
as a local	98.3	81.4	1.21			
as a dual	40.0	35.4	1.13	41.3	34.5	1.20
as a broker				9.6	7.9	1.21
#Transactions	450.3	345.9	1.30	145.7	112.3	1.30
through local	353.3	264.0	1.34			
through dual	96.9	81.8	1.18	117.4	90.9	1.29
through broker				28.2	21.4	1.32
Trade Size (in #contracts)	10.9	10.2	1.07	17.5	16.1	1.09
through local	12.0	11.3	1.06			
through dual	6.9	6.5	1.07	16.8	15.2	1.11
through broker				20.6	19.6	1.05
Bid-Ask Spread <sup>a</sup> (in \$)				6.4	5.6	1.14
through dual				6.7	5.9	1.13
through broker				4.3	3.4	1.26

<sup>a</sup> Bid-ask spread estimated as difference between the average (volume-weighted) customer buy and sell price.

Panel B: Signed Customer Volume (CTI4) Statistics (five min avg, 8:30-8:45)						
			Nonfarm, PPI, and CPI		Nonfarm Payroll Emp. Nonann	
			All Ann			
Customer Flow (1,000 contracts)	Mean	All	0.142**	0.301**	0.213	0.058*
		Dual	0.108**	0.224**	0.122	0.010
		Broker	0.034	0.077	0.091	0.048**
	St Dev	All	1.282	1.658	1.789	0.740
		Dual	1.164	1.477	1.622	0.663
		Broker	0.648	0.888	0.872	0.413

\*/\*\* indicates mean significant different from zero at the 95%/99% level.

flow. We find that net flow standard deviation on announcement days is higher than nonannouncement days and increases with the importance of the announcement.

## 2.3 Customer Order Informativeness

In this section, we pursue our main objective, which is to establish the increased informativeness of customer flow after a macro announcement. We consider this an important result, as it shows that intermediaries need off-exchange customer response to fully appreciate the effect of the macro announcement on the 30Y riskfree rate. Green (2004) documents empirical support as he finds an increased correlation between treasury returns and signed volume in the 15 minutes after an announcement. He cannot, however, identify that this information is in customer flow, as his signed volume is based on interdealer flow which, in addition to customer flow, also contains trades initiated by potentially superiorly informed intermediaries. In addition, the correlation might be endogenously biased upwards due to flow-based speculation by dual traders (see Appendix 2A).

### 2.3.1 Five-Minute Price Change Regressions on Customer Flow

We assess customer flow informativeness through a regression of five-minute price changes on aggregate signed customer flow. We prefer time-interval return regressions (as in Brandt and Kavajecz (2004) and Pasquariello and Vega (2007)) to trade return regressions (as in Green (2004)), as the aggregation alleviates any effect any time-stamp errors might have. Consistent with previous studies, we add the macro surprise to the regression and estimate:

$$p_{t,h} - p_{t,h-1} = d_a(\alpha_a + \beta_a\omega_{t,h}) + d_n(\alpha_n + \beta_n\omega_{t,h}) + \sum_k \gamma_k I_{k,t} S_{k,t} + \varepsilon_{t,h} \quad (2.2)$$

where  $p_{t,h}$  is 100 times the log price (to get % returns) at day  $t$  and five-minute interval  $h$ ,  $d_a$  ( $d_n$ ) is a dummy that is one on an announcement (nonannouncement) day, zero otherwise,  $\omega_{t,h}$  is the aggregate signed customer volume,  $S_{k,t}$  is the announcement surprise (see equation (2.1)),  $I_{k,t}$  is a dummy that is one for the time interval immediately after the announcement, zero otherwise,<sup>24</sup> and  $\varepsilon_{t,k}$  is the error term. The regression implicitly controls for feedback trading through inclusion of the macro surprise, i.e. any effect of  $\omega_{k,t}$  is identified off of the orthogonalized component relative to the other explanatory

---

<sup>24</sup>In the implementation, we do allow for the surprise to also affect later time intervals (i.e. 8:35-8:40, 8:40-8:45, ...) and find no significance. For robustness, we nevertheless repeat all analysis based on equation (2.2) and find that our results are not affected.

variables.<sup>25</sup> We emphasize that this is a contribution of our approach, as Green (2004, p.1210) only includes the surprise in the first transaction return after the announcement and therefore orthogonalizes only the first postannouncement transaction.

### Figure 2.3: Sensitivity of Treasury Return to Signed Customer Volume (CTI4)

This figure depicts the coefficient of signed customer volume (CTI4) in the 30Y treasury future return regressions. It plots this coefficient based on the estimation results of the following regression for all 15 minute intervals in the day:

$$p_{t,h} - p_{t,h-1} = d_a(\alpha_a + \beta_a \omega_{t,h}) + d_n(\alpha_n + \beta_n \omega_{t,h}) + \sum_k \gamma_k I_{k,t} S_{k,t} + \varepsilon_{t,h}$$

where  $p_{t,h}$  is 100 times the log price of the 30Y treasury futures at day  $t$  and five minute interval  $h$ ,  $d_a$  ( $d_n$ ) is a dummy that is one on an announcement (nonannouncement) day, zero otherwise,  $\omega_{t,h}$  is the aggregate signed customer volume (CTI4),  $S_{k,t}$  is the announcement surprise,  $I_{k,t}$  is a dummy that is one for the time interval immediately after the announcement, zero otherwise, and  $\varepsilon_{t,k}$  is the error term. For estimation, we use the Feasible Efficient GMM procedure with a Newey-West estimator (using three lags) for standard errors. The solid (dashed) line depicts the intraday pattern of  $\beta$  for announcement (nonannouncement) days; the vertical line represents the 8:30-8:45 announcement interval. A closed circle indicates a significant difference between announcement and nonannouncement days at the 99% level.

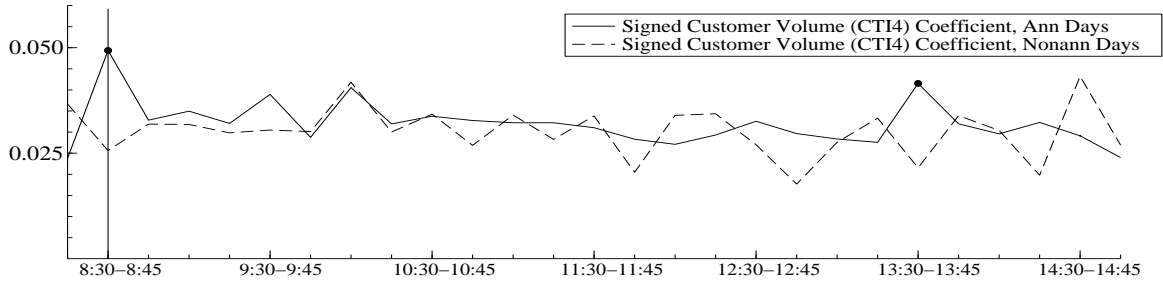


Figure 2.3 depicts the intraday pattern of customer flow informativeness on announcement as well as nonannouncement days. We estimate equation (2.2) separately for all 15-minute intervals in the trading day and test whether customer flow informativeness ( $\beta$ ) is significantly different on announcement days relative to nonannouncement days.<sup>26</sup> We find it to be significantly higher in the 15 minutes subsequent to the announcement and generally insignificant for the remainder of the day. Economically, informativeness roughly doubles in these 15 minutes and the intraday pattern is therefore comparable—in shape and magnitude—to the bid-ask spread pattern.

<sup>25</sup>This relies on one of the statistical properties of linear regression, which is that any multivariate regression coefficient can be obtained through univariate regression of the orthogonalized dependent variable on the orthogonalized explanatory variable, where the orthogonalization is with respect to the other regressors.

<sup>26</sup>For clarity, the  $d_a$  dummy of equation (2.2) is one for all five-minute intervals in a 15 minute period.

**Table 2.3: Regressions of 30Y Treasury Return on Signed Customer Volume**

This table reports the estimation results of the following regression:

$$p_{t,h} - p_{t,h-1} = d_a(\alpha_a + \beta_a \omega_{t,h}) + d_n(\alpha_n + \beta_n \omega_{t,h}) + \sum_k \gamma_k I_{k,t} S_{k,t} + \varepsilon_{t,h}$$

where  $p_{t,h}$  is 100 times the log price of the 30Y treasury futures at day  $t$  and five minute interval  $h$ ,  $d_a$  ( $d_n$ ) is a dummy that is one on an announcement (nonannouncement) day, zero otherwise,  $\omega_{t,h}$  is the aggregate signed customer volume (CTI4, *Customer Flow*) divided by 1,000,  $S_{k,t}$  is the announcement surprise,  $I_{k,t}$  is a dummy that is one for the time interval immediately after the announcement, zero otherwise, and  $\varepsilon_{t,k}$  is the error term. For estimation, we use the Feasible Efficient GMM procedure with a Newey-West estimator (using three lags) for standard errors. Panel A reports the estimates of the intercept and signed customer volume coefficients estimated for 8:30-8:45 based on five minute intervals and tests for equality of signed customer volume coefficients. In Panel B we split the estimated signed customer volume coefficients for the sample of all announcement days in three groups based on dispersion of beliefs. In particular, we follow Pasquariello and Vega (2007) and estimate the coefficients for days with high, medium and low dispersion of beliefs. High (low) dispersion is defined as the monthly forecasts' standard deviation to be in the top 70th (bottom 30th) percentile of its empirical distribution. The monthly forecasts' standard deviation is based on the standard deviation of forecasts for all available 8:30 announcements. We report  $t$ -values below coefficient estimates.

Panel A: 30Y Treasury Return Regressions					
			All Ann	Nonfarm, PPI, and CPI	Nonfarm Payroll Emp.
Customer Flow	Ann	$\beta_a$	0.0493** 10.4	0.0544** 6.06	0.0571** 3.77
	Nonann	$\beta_n$	0.0256** 9.67	0.0256** 9.67	0.0256** 9.67
Intercept	Ann	$\alpha_a$	-0.0118** -2.75	-0.0364** -2.69	-0.0974** -3.26
	Nonann	$\alpha_n$	0.0033* 2.32	0.0033* 2.32	0.0033* 2.32
#Observations	Total		2,181	1,365	1,152
	Ann		1,131	315	102
	Nonann		1,050	1,050	1,050
$R^2$			0.366	0.354	0.369
$p$ -value of $H_0: \beta_a = \beta_n$			0.0000**	0.0021**	0.0410*

\*/\*\* indicates significance at the 95%/99% level.

(continued on next page)



(continued from previous page)

				All Ann	Nonfarm, PPI, and CPI	Nonfarm Payroll Emp.	
Customer Flow	Ann	High Het	$\beta_{a,H}$	0.0596** 7.14	0.0677** 4.41	0.0643* 2.19	
		Med Het	$\beta_{a,M}$	0.0489** 7.32	0.0611** 4.5	0.0742** 2.99	
		Low Het	$\beta_{a,L}$	0.0365** 3.26	0.0290 1.53	0.0169 0.443	
	Nonann	High Het	$\beta_{n,H}$	0.0285** 6.34	0.0285** 6.34	0.0285** 6.34	
		Med Het	$\beta_{n,M}$	0.0303** 7.96	0.0303** 7.96	0.0303** 7.96	
		Low Het	$\beta_{n,L}$	0.0163** 3.04	0.0163** 3.04	0.0163** 3.04	
Intercept	Ann		$\alpha_a$	-0.0119** -2.8	-0.0363** -2.68	-0.0948** -3.28	
	Nonann		$\alpha_n$	0.0033* 2.27	0.0033* 2.27	0.0033* 2.27	
#Observations	Total			2,181	1,365	1,152	
	Ann			1,131	315	102	
	Nonann			1,050	1,050	1,050	
$R^2$				0.369	0.361	0.378	
$p$ -value of $H_0$ :				$\beta_{a,H} = \beta_{n,H}$	0.0010**	0.0142*	0.2270
				$\beta_{a,M} = \beta_{n,M}$	0.0157*	0.0289*	0.0801
				$\beta_{a,L} = \beta_{n,L}$	0.1040	0.5190	0.9880
				$\beta_{a,d} = \beta_{n,d}, d = H, M, L$	0.0002**	0.0102*	0.1670

\*/\*\* indicates significance at the 95%/99% level.

Panel A of Table 2.3 presents the results of a regression of returns on customer flow in the 15 minutes after the announcement (equation (2.2)). For the macro surprise coefficients ( $\gamma_k$ ), reported separately in Table 2.B2, we find that 9 out of the 15 announcement surprises significantly affects subsequent returns, where, generally, procyclical announcements (e.g., nonfarm payroll employment) negatively affect returns and countercyclical announcements (e.g., initial unemployment claims) positively affect returns. Among these announcements, we find that nonfarm payroll employment, producer price index (PPI), and consumer price index (CPI) have the largest economic impact. We therefore repeat all regressions with only these three announcement days and with only nonfarm payroll announcement days to verify that any effect we find increases with the importance of the news. Panel A of Table 2.3 shows that this is indeed the case for the customer flow informativeness differential across announcement and nonannouncement days.

We decompose the explained price change variance and find that 24.0% is due to customer order flow where we control for feedback trading. The R-squared shows that the announcement day regression explains 36.6% of price change variance. We use a Cholesky

decomposition on the explained part to judge how much is due to the immediate response to the announcement surprise and how much is due to subsequent customer flow. In the ordering, we choose to put the announcement surprise first so that effectively the contribution of customer flow is *net* of the component correlated with the announcement surprise. That is, mathematically, the effect it assigns to customer flow is based on customer flow orthogonalized relative to the surprise. The decomposition assigns 76.0% to the immediate response and 24.0% to (orthogonalized) customer flow. In the procedure, we find that 6.7% of the explanatory power of customer flow is effectively due to feedback trading as this is the size of the part that correlates with the announcement surprise.<sup>27</sup> The economic significance of customer flow is further demonstrated by the result that a one standard deviation increase in net customer flow on announcement days (see Table 2.2) causes the 30Y treasury return to be  $1.282 \times 0.0439 \times 100 = 6.3$  basispoints higher, which is substantial relative to a 23 basispoints volatility for the *full* 15-minute return (see Figure 2.2).

Panel B of Table 2.3 finds that customer flow informativeness increases with the dispersion in analyst forecasts. Pasquariello and Vega (2007) find theoretically as well as empirically that correlation between order flow and yield changes should increase with the dispersion of beliefs among market participants. We follow their empirical approach and interact our customer flow variable with dummy variables that differentiates months with low, medium, and high analyst forecast dispersion. We find that, as expected, customer flow informativeness increases monotonically with dispersion. In our econometric tests on the coefficient differential across announcement and nonannouncement days, we find most significance for the high dispersion months, which is not surprising. In the joint test on the differential across all three forecast dispersion regimes, we find a significant differential only for the first two categories: (i) all announcements and (ii) the Nonfarm, PPI, and CPI announcements. It seems that we lack statistical power to also reject the null of no differential for the Nonfarm only announcements.

### 2.3.2 An Alternative Interpretation of the Regression Coefficient

So far, we interpret our regression coefficient as trade informativeness. In inactive markets, part of the price change correlated with order flow is transitory in nature in order to compensate a liquidity supplier for the cost of market-making. We consider such effect

---

<sup>27</sup>We decompose the variation of  $X'\beta$  where  $X$  is the matrix of explanatory variables and  $\beta$  is vector of coefficient estimates. The customer flow is the last element in the  $X$ . Cholesky decomposes the customer flow (explanatory) variation into a part that is projected onto the macro surprises (“feedback part”, 6.7%) and an orthogonalized part (93.3%). In the procedure we subtract the explained variation on nonannouncement days to single out the effect due to the *increased* informativeness.

unlikely for our five-minute regressions in what is a very active market; we find 172.9 intermediaries active who collectively generate 595.9 transactions in the average five-minute interval on announcement days (see Table 2.2).

We rerun the regressions with decomposed customer flow to provide further evidence of informativeness. We decompose customer flow according to whether it reaches the floor through brokers (who do not trade for own-account that day) or through duals. The results in Table 2.4 show that yield changes are only significantly more sensitive to dual-intermediated customer flow on announcement days.<sup>28</sup> The unchanged sensitivity to broker-intermediated customer flow is not a straightforward result of order size, as, if anything, brokers intermediate larger customer orders than duals do (see Table 2.2). Thus, this differential in sensitivity across dual- and broker-intermediated flow rules out a noninformation explanation based on increased price concession due to market-making costs, as this would affect all customer flow equally. Rather, we believe that the intermediary's decision to trade for own account is endogenous and depends on whether she traces informed customers in her customer flow in the aftermath of the announcement. We note that this result is consistent with the higher bid-ask spread reported for dual-intermediated customer trades relative to broker-intermediated customer trades (see Section 2.2.3).

## 2.4 Intermediary's Own-Account Trading

With the result of increased customer flow informativeness after an information event, we analyze whether intermediaries benefit from direct access to customer flow through own-account trading. As mentioned in the introduction, screening out the customer flow and discriminating informed from uninformed customers, the intermediary's rational strategy is to trade along with informed customers and opposite to uninformed orders (see Appendix 2A).

### 2.4.1 Is Direct Access to Customer Flow Profitable?

Inspired by Fishman and Longstaff (1992), we analyze own-account trading profitability for intermediaries with access to customer flow (duals) and intermediaries without such

---

<sup>28</sup>We admit that the lack of significance for broker-intermediated flow might be the result of statistical power as in the average five-minute interval we find that the ratio of the number of active duals to the number of active brokers is roughly four (see Table 2.2). On the other hand, the sign of the differential of broker-intermediated customer flow across announcement days is wrong in two of the three cases.

**Table 2.4: Dual- vs. Broker-Intermediated Signed Customer Volume (CTI4)**

This table follows up on Table 2.3 and decomposes signed customer volume (CTI4) into dual- vs. broker-intermediated signed customer volume. It reports the estimation results of the following regression:

$$p_{t,h} - p_{t,h-1} = d_a(\alpha_a + \beta_a^d \omega_{t,h}^d + \beta_a^b \omega_{t,h}^b) + d_n(\alpha_n + \beta_n^d \omega_{t,h}^d + \beta_n^b \omega_{t,h}^b) + \sum_k \gamma_k I_{k,t} S_{k,t} + \varepsilon_{t,h}$$

where  $p_{t,h}$  is 100 times the log price of the 30Y treasury futures at day  $t$  and five minute interval  $h$ ,  $d_a$  ( $d_n$ ) is a dummy that is one on an announcement (nonannouncement) day, zero otherwise,  $\omega_{t,h}^d$  ( $\omega_{t,h}^b$ ) is the aggregate signed customer volume (CTI4, *Customer Flow*) intermediated by duals (brokers) divided by 1,000,  $S_{k,t}$  is the announcement surprise,  $I_{k,t}$  is a dummy that is one for the time interval immediately after the announcement, zero otherwise, and  $\varepsilon_{t,k}$  is the error term. For estimation, we use the Feasible Efficient GMM procedure with a Newey-West estimator (using three lags) for standard errors. We report the estimates of the intercept and signed customer volume coefficients estimated for 8:30-8:45 based on five minute intervals and tests for equality of signed customer volume coefficients. We report  $t$ -values below coefficient estimates.

Return Regressions, Dual- vs. Broker-Intermediated Signed Customer Volume (CTI4)						
				All Ann	Nonfarm, PPI, and CPI	Nonfarm Payroll Emp.
Customer Flow	Ann	Dual	$\beta_a^d$	0.0562** 10.0	0.0650** 5.70	0.0762** 3.60
		Broker	$\beta_a^b$	0.0238* 2.53	0.0209 1.27	-0.0197 -0.494
	Nonann	Dual	$\beta_n^d$	0.0265** 8.92	0.0265** 8.92	0.0265** 8.92
		Broker	$\beta_n^b$	0.0230** 4.77	0.0230** 4.77	0.0230** 4.77
Intercept	Ann		$\alpha_a$	-0.0119** -2.79	-0.0367** -2.72	-0.0931** -3.23
	Nonann		$\alpha_n$	0.0035* 2.34	0.0035* 2.34	0.0035* 2.34
#Observations	Total			2,181	1,365	1,152
	Ann			1,131	315	102
	Nonann			1,050	1,050	1,050
$R^2$				0.375	0.366	0.396
$p$ -value of $H_0$ :			$\beta_a^d = \beta_n^d$	0.0000**	0.0011**	0.0202*
			$\beta_a^b = \beta_n^b$	0.9400	0.9000	0.2880
			$\beta_a^d = \beta_a^b$	0.0036**	0.0343*	0.0628
			$\beta_n^d = \beta_n^b$	0.5110	0.5110	0.5110

\*/\*\* indicates significance at the 95%/99% level.

access (locals) in the 15 minutes after the announcement. We define profitability as:

$$\pi_{kt} = \frac{\left( \sum_{j=1}^{N_{kt}^s} q_{jkt}^s P_{jkt}^s - \sum_{j=1}^{N_{kt}^b} q_{jkt}^b P_{jkt}^b + (\sum_{j=1}^{N_{kt}^b} q_{jkt}^b - \sum_{j=1}^{N_{kt}^s} q_{jkt}^s) REF P_t \right)}{\max(\sum_{j=1}^{N_{kt}^b} q_{jkt}^b, \sum_{j=1}^{N_{kt}^s} q_{jkt}^s)} \quad (2.3)$$

where  $\pi_{kt}$  is the profit per round-trip contract<sup>29</sup> for intermediary  $k$  on day  $t$ ,  $N_{kt}^b$  ( $N_{kt}^s$ ) is the total number of buys (sells),  $q_{jkt}^b$  ( $q_{jkt}^s$ ) is the quantity of the  $j$ th transaction in terms of number of contracts,  $P_{jkt}^b$  ( $P_{jkt}^s$ ) is the associated price, and  $REF P_t$  is the reference price in day  $t$ . The profit calculation assumes that the intermediary starts with zero inventory and liquidates his end-of-period position at a reference price  $REF P_t$ . We present results where we set the reference price equal to the last transaction price in the measurement interval. For robustness, we also analyze profits based on the end-of-day settlement price as reference price, which gives qualitatively similar results.<sup>30</sup> We note that, by construction, this profit is *net* of adverse-selection cost (as it aggregates across multiple subsequent transactions and therefore includes losses due to adverse selection), but *gross* of market-making cost (e.g., inventory cost, order-processing cost).

Table 2.5 reports round-trip profitability per contract for duals and locals on announcement and nonannouncement days. We find very large standard deviations due to some extreme positive and negative observations. We therefore prefer a nonparametric test on median differences to the standard test on mean differences (see also Fishman and Longstaff (1992)). A \* (\*\*) indicates a significant difference between announcement and nonannouncement days at the 95% (99%) level, whereas x (xx) indicates a significant difference between dual profitability and local profitability. We emphasize two important results.

First, we find that a local's profitability on own-account trading is higher on announcement days and increases with the importance of the announcement. We find that locals make a median \$0.0 per contract traded round-trip on nonannouncement days. It is significantly higher on announcement days, \$7.8, which amounts to an approximate<sup>31</sup> \$1,063 per local for the full 15 minutes. It further increases with the importance of the announcement to \$14.8 per contract on nonfarm, PPI, and CPI days to \$23.7 on the nonfarm days. We interpret this as evidence of increased profits to compensate for the higher cost of carrying inventory through volatile times.

---

<sup>29</sup>We use a per-contract profit measure to control for trade activity, as locals are more active than duals.

<sup>30</sup>These results are not included here but are available upon request.

<sup>31</sup>Based on 264.0 ("single-trip") transactions of 11.3 contracts by 81.4 locals per five minutes on nonannouncement days, a volume increase of 300% in the 15 minutes subsequent to an announcement, a 21% increase active locals and a negligible increase in trade size on announcement days, i.e.  $\$1.063 \approx \$7.8 * 264 * .5 * 11.3 * 3 * 3 / (81.4 * 1.21)$  (see Table 2.2 and Figure 2.2).

**Table 2.5: Own-Account Trading Profits by Trader Type**

This table reports summary statistics on the cross-sectional distribution of proprietary trading profits in the 8:30-8:45 interval by trader type. We distinguish two types: those who also trade for customers on the same day, i.e. duals, and those who do not trade for customers on that day, i.e. locals. We follow Fishman and Longstaff (1992) and calculate the profits per contract traded round trip. That is, for each trader we subtract the value of purchases from the value of sales and add the value of end-of-period inventory (assuming zero inventory at the start). We divide this by the total number of contracts traded to arrive at a profit per contract traded round trip. Formally, we calculate:

$$\pi_{kt} = \left( \sum_{j=1}^{N_{kt}^s} q_{jkt}^s P_{jkt}^s - \sum_{j=1}^{N_{kt}^b} q_{jkt}^b P_{jkt}^b + \left( \sum_{j=1}^{N_{kt}^b} q_{jkt}^b - \sum_{j=1}^{N_{kt}^s} q_{jkt}^s \right) REF P_t \right) / \max \left( \sum_{j=1}^{N_{kt}^b} q_{jkt}^b, \sum_{j=1}^{N_{kt}^s} q_{jkt}^s \right),$$

where  $\pi_{kt}$  is the profit per round-trip contract for intermediary  $k$  on day  $t$ ,  $N_{kt}^b$  ( $N_{kt}^s$ ) is the total number of buys (sells),  $q_{jkt}^b$  ( $q_{jkt}^s$ ) is the quantity of the  $j$ th transaction in terms of number of contracts,  $P_{jkt}^b$  ( $P_{jkt}^s$ ) is the associated price, and  $REF P_t$  is the reference price in day  $t$ . We assume any remaining inventory is valued at the last price before 8:45, thus  $REF P_t$  is the last observed price before 8:45. We show the mean, standard deviation (*St Dev*) and the three quartiles (*25% Quant*, *Median* and *75% Quant*) of the cross-sectional distribution (across intermediaries) of own-account trading profits (with the number of trader days in each group in the column *#Trader Days*).

Own-Account Trading Profits per Contract Traded Round Trip						
	#Trader Days	Mean	St Dev	25% Quant	Median	75% Quant
<b>Locals</b>						
nonannouncement days	64,713	2.5	38.2	-13.5	0.0 <sup>xx</sup>	20.8
all announcement days	83,516	8.4	67.4	-13.2	7.8 <sup>**</sup> , <sup>xx</sup>	31.2
nonfarm, PPI, and CPI	25,301	17.0	93.0	-12.1	14.8 <sup>**</sup> , <sup>xx</sup>	43.9
nonfarm payroll emp.	8,242	26.7	117.8	-11.1	23.7 <sup>**</sup> , <sup>xx</sup>	62.5
<b>Duals</b>						
nonannouncement days	17,181	4.6	46.7	-15.6	2.2 <sup>xx</sup>	31.2
all announcement days	26,474	16.5	99.0	-13.4	13.9 <sup>**</sup> , <sup>xx</sup>	40.5
nonfarm, PPI, and CPI	8,381	29.6	142.2	-14.2	22.8 <sup>**</sup> , <sup>xx</sup>	62.5
nonfarm payroll emp.	2,709	49.0	199.1	-12.5	31.3 <sup>**</sup> , <sup>xx</sup>	101.6

<sup>\*\*</sup>/<sup>\*\*</sup> indicates significance relative to nonannouncement days at the 95%/99% level.

<sup>x</sup>/<sup>x</sup> indicates significance relative to the other trader type at the 95%/99% level (i.e. a comparison across local and dual profit).

Second, we find that duals appear to benefit from direct access to customer flow as they trade more profitably for own-account than locals do and, more importantly, this differential is higher on announcement days. We find that duals make a median \$2.2 per contract on nonannouncement days, which is significantly higher than the \$0.0 locals make. The result indicates that customer order flow is informative even on nonannouncement days. The important result, however, is that this differential is significantly higher on announcement days. Round-trip profit per contract is \$6.1 (\$13.9-\$7.8) higher for duals on announcement days, \$8.0 higher on nonfarm, PPI, and CPI days, and \$7.6 higher on nonfarm days.

### 2.4.2 The Alternative Explanation: Superior Trading Skills

Fishman and Longstaff (1992) entertain the alternative explanation that some traders have superior trading skill—trade more profitably for own-account—and customers choose to trade through these intermediaries to benefit from their skill. Thus, the correlation we document between trading for customers and own-account profitability might be spurious. To control for skill, Fishman and Longstaff (1992, Table 4) analyze trading profit of “nonpure” duals, i.e. intermediaries who some days trade for own-account only (local days) and other days trade both for own-account and for customers (dual days). We use the same approach in our sample.

Panel A of Table 2.6 reports the profit differential of nonpure duals on the days they have access to customer flow relative to the days that they do not have access (i.e. own-account profitability on dual days minus own-account profitability on local days). We find that, on their dual days, they earn a significantly higher profit than on their local days—the median differential is \$5.6 per round-trip contract. We then separate announcement and nonannouncement days and do the same analysis. Interestingly, we find a significantly increased profit for announcement days only. For nonannouncement days, we find no statistical difference, consistent with Fishman and Longstaff (1992). We conclude that, after control for trading skill, we continue to find support for the premise that intermediaries rely on off-exchange customer flow to fully appreciate the effect of macro news and benefit from discriminating the informed traders in their customer flow.

Panel B compares a nonpure dual’s local-day profit to a (pure) local’s profit and finds no evidence of superior trading skill. For nonannouncement days, we find a profitability of \$0.0 and \$0.1 on local days of nonpure duals and locals, respectively, and the difference is not statistically significant. For announcement days, we find a similar results as the profitability is \$7.8 and \$7.8, respectively, where again the difference is not significant. This higher profitability on announcement days is likely to reflect increased cost of market-making, e.g. due to higher inventory costs as result of higher volatility (see Fig-

**Table 2.6: Own-Account Trading Profits of Nonpure Duals and Pure Locals**

This table reports own-account trading profits per round trip in the 8:30-8:45 interval of nonpure duals, i.e. intermediaries who have both dual days (i.e. days they also trade for customers) and local days (i.e. days they do not trade for customers). Panel A reports cross-sectional statistics across all nonpure duals on the difference in average own-account profit for dual days and local days. Panel B reports cross-sectional statistics for the average own-account profit of nonpure duals on local days and similar statistics for the average own-account profit of pure locals (i.e. intermediaries that never trade for customers). To obtain own-account trading profits for each trader we subtract the value of purchases from the value of sales and add the value of end-of-period inventory (assuming zero inventory at the start). We divide this by the total number of contracts traded to arrive at a profit per contract traded round trip.

Panel A: Nonpure Duals' Profit Advantage on their Dual Days relative to their Local Days			
Difference in Profits	All Days	Nonann Days	Ann Days
#Nonpure Duals	234	184	200
Mean Profit Advantage	8.6	2.8	13.5
Standard Deviation	59.1	35.9	68.0
25% Quantile	-8.3	-12.6	-10.2
Median	5.6	3.5	5.0
75% Quantile	24.6	14.7	34.7
%-age Coeff's positive	63.2	55.4	58.0
Test $z$ -statistic <sup>a</sup>	4.05	1.47	2.26

<sup>a</sup> Test statistic standard normal under  $H_0$ .

Panel B: Trading Profits on Local Days, Pure Locals vs Nonpure Duals						
	#Trader Days	Mean	St Dev	25% Quant	75% Median Quant	75% Quant
Local Days of Pure Local						
nonannouncement days	33,083	2.8	37.4	-12.7	0.1	20.4
all announcement days	42,808	8.7	67.2	-13.6	7.8**	31.2
nonfarm, PPI, and CPI	18,499	16.7	90.2	-11.7	14.4**	42.7
nonfarm payroll emp.	6,911	26.5	115.8	-11.6	23.4**	61.7
Local Days of NonPure Dual						
nonannouncement days	27,880	2.3	38.3	-13.9	0.0	20.8
all announcement days	36,061	8.5	67.4	-12.9	7.8**	31.2
nonfarm, PPI, and CPI	5,887	17.8	101.6	-13.5	15.6**	47.8
nonfarm payroll emp.	1,100	27.7	131.2	-7.8	25.2**	65.4

\*/\*\* indicates significance relative to nonannouncement days at the 95%/99% level.

$x/x$  indicates significance relative to the other trader type at the 95%/99% level (i.e. a comparison across local days of pure local and local days of nonpure dual profit).



ure 2.2). In sum, the insignificant difference between nonpure duals and locals suggests that idiosyncratic trading skill is not important in explaining cross-sectional differences in own-account trading profitability. Hence, this makes the alternative explanation for our results unlikely.

### 2.4.3 Do Profits Increase with the Level of Customer Flow Access?

We exploit the cross-section of duals to further establish a relationship between own-account profitability and access to customer order flow. We regress a dual trader's profit per contract in the 15 minutes of postannouncement trading on a measure of access to customer flow and various control variables:

$$\pi_{l,t} = \alpha + \beta_1 CUST_{l,t} + \beta_2 VOLA_t + \beta_3 COMP_t + \sum_k \gamma_k |S_{k,t}| + \varepsilon_{l,t} \quad (2.4)$$

where  $\pi_{l,t}$  is the profit per contract traded round trip of dual trader  $l$  in the 15 minutes of postannouncement trading on day  $t$ ,  $CUST_{l,t}$  proxies for her access to customer flow in these 15 minutes (e.g., number of customer trades executed per contract traded round trip),  $VOLA_t$  is our volatility measure (see Section 2.2.3),  $COMP_t$  is a competition proxy and is defined as the ratio of the number of active intermediaries who trade for customers (i.e., dual and brokers) and the number of customer trades,  $S_{kt}$  is the macro surprise of announcement type  $k$ , and  $\varepsilon_{lt}$  is the error term.<sup>32</sup> We control for a potential competition effect, as Wahal (1997), for example, finds that the number of dealers matters for the bid-ask spread in the NASDAQ market, which he interprets to be “consistent with the competitive model of dealer pricing”. We relate a dual trader's profit to her access to customer flow and, therefore, build a competition proxy on how many rivals she has for each customer trade (i.e., duals and brokers). In addition to equation (2.4), we perform a regression where we replace all control variables by a day dummy to kill all the time effect and we therefore only get traction from the cross-section. This makes it generally harder to find a significant estimate of  $\beta_1$ .<sup>33</sup>

---

<sup>32</sup>We use the indicator  $l$  to relabel duals every day to minimize notational burden.

<sup>33</sup>The model with controls is nested in the time dummy model, as the controls are effectively spanned by the time dummies, i.e. they are a linear combination of these dummies.

**Table 2.7: Determinants of Dual Trader's Own Account Profits on Announcement Days**

This table reports the estimation results of the following regression:

$$\pi_{lt} = \alpha + \beta_1 CUST_{lt} + \beta_2 VOLA_t + \beta_3 COMP_t + \sum_k \gamma_k |S_{kt}| + \varepsilon_{lt}$$

where  $\pi_{lt}$  is dual  $l$ 's own-account profit per round trip trade in the 15 minutes following the announcement on day  $t$ ,  $CUST_{lt}$  proxies for dual trader  $l$ 's access to customer flow,  $VOLA_t$  is the volatility measure,  $COMP_t$  is a competition proxy and is defined as the ratio of the number of active intermediaries who trade for customers (i.e. dual and brokers) and the number of customer trades,  $S_{kt}$  is the macro surprise of announcement type  $k$ , and  $\varepsilon_{lt}$  is the error term. We use four proxies for a dual's access to customer flow (CTI4): the number of trades of dual  $l$  on day  $t$  that come from customers  $\sum_j |D_{j,l,t}^c|$  (model (1)), the absolute value of the sum of the signed number of customer trades  $|\sum_j D_{j,l,t}^c|$  (model (2)), the total volume of dual  $l$  on day  $t$  that comes from customers  $\sum_j q_{j,l,t}^c$  (model (3)) and the absolute value of the sum of the signed volume  $|\sum_j D_{j,l,t}^c q_{j,l,t}^c|$  (model (4)) where  $D_{j,l,t}^c$  represents the direction (+1 for buy, -1 for sell) of trade  $j$  for trader  $l$  on day  $t$ . We scale the proxies for access to customer flow  $CUST_{lt}$  with the number of round trips ( $\#RndTrips_{l,t}$ ) for each dual  $l$  on each day  $t$  as this is also done for our profits measure  $\pi_{lt}$ . All regressors are demeaned to let the intercept represent the average trading profit per round trip in the 8:30-8:45 interval of a dual on an announcement day. For estimation, we use the Feasible Efficient GMM procedure with a Newey-West estimator (using three lags) for standard errors.

Dependent Variable: Dual's Trading Profit per Contract Traded Round Trip in the 8:30-8:45 interval on Ann Days												
	(1)	(1')	(1'')	(2)	(2')	(2'')	(3)	(3')	(3'')	(4)	(4')	(4'')
Proxies for $CUST_{lt}$												
customer trades												
$\sum_j  D_{j,l,t}^c /\#RndTrips_{l,t}$	1.07** 3.1	0.646 1.9	0.600 1.75									
signed customer trades												
$ \sum_j D_{j,l,t}^c /\#RndTrips_{l,t}$				3.14** 4.44	2.64** 3.82	2.76** 3.78						
customer volume												
$\sum_j q_{j,l,t}^c/\#RndTrips_{l,t}$							0.0218 1.59	0.00926 0.653	0.00744 0.514			
signed customer volume												
$ \sum_j D_{j,l,t}^c q_{j,l,t}^c /\#RndTrips_{l,t}$										0.0656* 2.16	0.0459 1.47	0.0444 1.45
Intercept	16.5** 26.1	16.5** 26.9		16.5** 26.1	16.5** 26.9		16.5** 26	16.5** 26.8		16.5** 26.1	16.5** 26.8	
Controls												
volatility		2.93** 3.18			2.87** 3.12			3.00** 3.24			3.00** 3.23	
competition		-28.0* -2.22			-28.0* -2.23			-30.5* -2.44			-30.2* -2.42	
surprise?		yes			yes			yes			yes	
time dummy?			yes			yes			yes			yes
#Observations	26,474	26,474	26,474	26,474	26,474	26,474	26,474	26,474	26,474	26,474	26,474	26,474
$R^2$	0.002	0.017	0.040	0.004	0.019	0.042	0.000	0.016	0.039	0.000	0.016	0.040

\*/\*\* indicates significance at the 95%/99% level.

Table 2.7 shows that a dual's own-account profitability increases with access to customer flow ( $\beta_1 > 0$ ), but we only find strong significance if we use the signed customer trades proxy. We number the regression results based on the four proxies we use: number of trades, sum of signed trades, volume, and sum of signed volume. We use trades as well as volume to account for a potential trade size effect and we use signed and unsigned to account for a potential imbalance effect. Per proxy, we perform three regressions: a univariate regression, a regression with controls, and a regression with day dummies. The results show a positive coefficient for access to customer flow across all regressions and customer flow proxies, but we only find robust statistical significance for the signed customer trade proxy. Economically, the effect is substantial, as a one standard deviation increase in the signed trades proxy (1.94) earns the intermediary an additional  $1.94 * \$2.76 = \$5.35$  per contract on her own-account trades, which is a 39% increase relative to her \$13.9 median profit on announcement days.<sup>34</sup> We further find, not surprisingly, that profits increase with volatility (e.g., to reflect more costly inventory keeping) and decrease with the level of competition.

The finding that the trade-based proxy shows stronger result than to the volume-based proxy is not surprising in view of the flow-based speculation argument. At first sight, the weak result on customer volume is counterintuitive as one expects large customer orders to be more informative than small ones. Order size, however, is not private information to the intermediary as she has to execute the customer order on the floor before trading for own-account. So size, even though potentially informative, is not private information to her. Customer identity on the other hand is not revealed in the trading process and remains private information to the intermediary. It is therefore not surprising that having access to a multiplicity of customer orders is a key driver of own-account profitability rather than access to large customer orders.

## 2.5 Who Pays for the Dual's Increased Profit?

The previous section documents that intermediaries with direct access to customer flow benefit through own-account trading. We interpret this result as evidence of flow-based speculation, where the intermediary privately observes the identity of the submitting customer. In Appendix 2A, we illustrate the mechanism in a simple extension of the Kyle (1985) model where the intermediary benefits from discriminating informed from uninformed customer flow. She trades in the same direction as her informed customer and opposite to her uninformed customer.

In the single intermediary world with a rational, zero-profit market maker, the interme-

---

<sup>34</sup>See Table 2.2 and Table 2.7.

diary's increased profit is paid for by her customers through an increased price concession that the market maker charges to protect herself against flow-based speculation. In an multiple intermediaries setting, who pays for the dual's increased profit critically depends on the extent that market makers can infer which intermediaries are likely to engage in flow-based speculation. The extremes are that (i) market makers get no signal or (ii) that they can fully discriminate the flow-based speculators. In the first case, they charge all intermediaries the same price concession and the dual's increased profit are effectively paid for by all customers. In the second case, the market maker only charges increased price concession to the flow-based speculators and, as a result, only their customers pay the increased profit.

### 2.5.1 Profitability of Dual- and Broker-Intermediated Customer Orders

We analyze customer profits in the 15 minutes after the announcement to study whether dual-intermediated customers pay a disproportionate part of their intermediary's increased profit. We calculate customer profits based on equation (2.3) where we replace the intermediary's own-account trades (CTI1) by customer trades (CTI4).

Table 2.8 shows that a dual's customer seems to pay a disproportionate part of her increased profit. We find that on announcement days the broker-intermediated customer trades earn a \$0.0 median profit per round-trip contract, whereas dual-intermediated profit is significantly lower and amounts to \$-7.3 per contract.<sup>35</sup> This difference remains for the narrower sets of important announcements, but it is insignificant probably due to low power as the sample is considerably smaller. Our results differ from Fishman and Longstaff (1992) who find significantly higher profit for dual-intermediated customer flow, which is consistent with a model where a market maker cannot infer which intermediary has the informed customer orders. That is, she charges all the same price concession, which should make customer profits higher for dual-intermediated customer flow as it contains the positive profit of the informed customer order. The broker-intermediated customer flow, on the other hand, does not contain such informed orders. We interpret our result as a sign that market makers do get a signal on who has the informed customer flow and rationally charge them an additional price concession (to protect themselves against flow-based speculation). This interpretation is consistent with our earlier result that the bid-ask spread is higher for dual-intermediated customer orders relative to broker-intermediated ones.

The negative customer profit result begs the question: Why do customers trade

---

<sup>35</sup>The negative sign is intuitive as it indicates that customers pay for demanding liquidity.

**Table 2.8: Customer Profits of Dual- vs. Broker-Intermediated Trades**

This table reports customer trading profits in the 8:30-8:45 interval of dual- and broker- intermediated customer trades, where dual traders also trade for own-account on that day and brokers do not. We follow Fishman and Longstaff (1992) and calculate the aggregate customer profits per contract traded round trip. That is, for each dual and broker trader we subtract the value of her customer purchases from the value of her customer sales and add the value of end-of-period inventory (assuming zero inventory at the start). We divide this by the total number of customer contracts traded to arrive at a profit per contract traded round trip. Formally, we calculate:

$$\pi_{kt} = \left( \sum_{j=1}^{N_{kt}^s} q_{jkt}^s P_{jkt}^s - \sum_{j=1}^{N_{kt}^b} q_{jkt}^b P_{jkt}^b + \left( \sum_{j=1}^{N_{kt}^b} q_{jkt}^b - \sum_{j=1}^{N_{kt}^s} q_{jkt}^s \right) REF P_t \right) / \max \left( \sum_{j=1}^{N_{kt}^b} q_{jkt}^b, \sum_{j=1}^{N_{kt}^s} q_{jkt}^s \right),$$

where  $\pi_{kt}$  is the customer profit per round-trip contract for intermediary  $k$  on day  $t$ ,  $N_{kt}^b$  ( $N_{kt}^s$ ) is the total number of customer buys (sells),  $q_{jkt}^b$  ( $q_{jkt}^s$ ) is the quantity of the  $j$ th customer transaction in terms of number of contracts,  $P_{jkt}^b$  ( $P_{jkt}^s$ ) is the associated price, and  $REF P_t$  is the reference price in day  $t$ . We assume any remaining inventory is valued at the last price before 8:45, thus  $REF P_t$  is the last observed price before 8:45. We show the mean, standard deviation (*St Dev*) and the three quartiles (*25% Quant*, *Median* and *75% Quant*) of the cross-sectional distribution (across intermediaries) of her customers' aggregate trading profits (with the number of trader days in each group in the column *#Trader Days*).

Customer Profits per Contract Traded Round Trip						
	#Trader Days	Mean	St Dev	25% Quant	Median	75% Quant
Dual-Intermediated Customer Trades						
nonannouncement days	17,181	-3.0	65.1	-32.5	0.0 <sup>x</sup>	31.3
all announcement days	26,474	-12.6	129.5	-67.7	-7.3 <sup>**</sup> , <sup>xx</sup>	49.0
nonfarm, PPI, and CPI	8,381	-22.7	175.8	-104.2	-17.5 <sup>**</sup>	63.5
nonfarm payroll emp.	2,709	-35.0	225.0	-147.1	-25.8 <sup>**</sup>	87.3
Broker-Intermediated Customer Trades						
nonannouncement days	6,567	-1.3	70.2	-31.3	0.0 <sup>x</sup>	31.3
all announcement days	9,034	-7.3	143.3	-62.5	0.0 <sup>**</sup> , <sup>xx</sup>	58.0
nonfarm, PPI, and CPI	2,843	-14.9	200.3	-101.7	-11.3 <sup>**</sup>	73.9
nonfarm payroll emp.	970	-23.2	250.9	-145.4	-19.3 <sup>**</sup>	94.1

<sup>\*\*</sup>/<sup>\*\*</sup> indicates significance relative to nonannouncement days at the 95%/99% level.

<sup>x</sup>/<sup>xx</sup> indicates significance relative to the other trader type at the 95%/99% level (i.e. a comparison across dual and broker aggregate customer profit).

through duals if they seem to lose money? One possible explanation is that customers do not know *ex-ante* if their intermediary will dual trade that day or not. Fishman and Longstaff (1992) emphasize that the intermediary's decision is endogenous and depends on whether she receives informed customer flow. It is easily imaginable that this is a hard to predict event as informed investors have an incentive to randomize across intermediaries to hide their type as much as they can.

## 2.6 Conclusion

We exploit a comprehensive dataset of 42.5 million transactions in the 1994-1997 30Y treasury futures market that captures 95% of overall volume (i.e. including the underlying). We are able to discriminate the off-exchange customer orders and find that they exhibit increased informativeness in the 15 minutes after an 8:30 macro announcement. This suggests that intermediaries rely on off-exchange customer orders to fully appreciate how macro news affects the 30Y riskfree rate. Green (2004) documents the increased informativeness for interdealer order flow; we contribute and show that an important channel is off-exchange customer "response" to the news. The market appears to aggregate micro information on imperfectly known preferences and endowments throughout the economy (see, e.g., Saar (2007)).

We generate further evidence for customer flow informativeness through an analysis of own-account profitability in a large cross-section of 3,382 intermediaries. That is, if (i) customer flow is informative and if (ii) there is heterogeneity in order informativeness across customers, then observing customer identity is private information to the intermediary. In our market she has to trade her customer order on the floor before being able to trade for own-account, but customer identity does not have to be revealed on the floor. She therefore continues to benefit from having observed customer identity information and can trade on it *ex-post* for own account. We find supportive evidence for such flow-based speculation. First, we find that in the 15 minutes after the announcement, intermediaries with access to customer flow trade significantly more profitable for own account than intermediaries without such access. The difference is a 78% higher median profit per contract. Second, among these intermediaries with access to customer flow, own-account profitability increases with access to signed customer trades.

Overall, our findings suggest that the trading process aggregates micro demand from off-exchange customers to discover macro variables such as the 30Y riskfree rate. This should give further empirical foundation to current macro models that build on agents' decisions at the micro level.

## Appendix 2A: Flow-based Speculation

In this appendix, we use the Kyle (1985) model to illustrate that price impact is increased in the presence of an intermediary who trades for her own account. The key engine for this result is that, contrary to the intermediary, the market maker does not observe the composition of customer flow. The intuition is that the rents earned by the intermediary are paid for by customers through an increased price impact (as the market maker earns zero rents).

Suppose  $v$ , the unknown payoff of the asset, is normally distributed with zero expectation and variance equal to  $\sigma_v^2$ . The customers consist of an informed investor who knows  $v$  and an uninformed investor who exogenously trades an amount  $u$ , which is normally distributed with zero expectation and variance  $\sigma_u^2$ .

Without an intermediary, Kyle (1985) finds the following unique linear equilibrium:

$$X(v) = \beta v, \beta = \frac{1}{2\lambda} \quad (\text{linear strategy of informed investor}) \quad (2.5)$$

$$P(\omega) = E[v|\omega] = \lambda\omega, \lambda = \frac{1}{2} \frac{\sigma_v}{\sigma_u} \quad (\text{market maker earns zero rents}) \quad (2.6)$$

where  $\omega = X(v) + u$  is the aggregate order flow the market maker receives.

We deviate from the standard setting and introduce an intermediary who observes the origination of the customer order and adds her own order  $y$  before submitting the aggregate order flow to the market maker. We restrict  $y$  to be a linear order:

$$y = \alpha v + \gamma u \quad (2.7)$$

The informed trader rationally anticipates the intermediary's action and internalizes her response when choosing  $\beta$ . We work backward and solve sequentially:

1. We condition on  $\lambda$  and  $\beta$  to maximize the intermediary's expected profit:

$$E[(P - v)y] = E[(\lambda\omega - v)y] = E[(\lambda((\alpha + \beta)v + (1 + \gamma)u) - v)(\alpha v + \gamma u)] \quad (2.8)$$

which yields:

$$\alpha = \frac{1}{2} \left( \frac{1}{\lambda} - \beta \right), \quad \gamma = -\frac{1}{2} \quad (2.9)$$

We find that (i) the intermediary trades less aggressively on the true value  $v$  if the informed customer submits a larger order (higher  $\beta$ ) or if liquidity is lower (higher  $\lambda$ ) and that (ii) she rationally takes the opposite side of the uninformed order ( $\gamma < 0$ ).

2. We condition on  $\lambda$  and on the intermediary's action to maximize profits for the

informed trader and find:

$$\beta = \frac{1}{2\lambda} \quad (2.10)$$

The result is that the aggregate order loads more heavily on the signal ( $\alpha + \beta = \frac{3}{4\lambda} > \frac{1}{2\lambda}$ ) (see equation (2.5)) as the informed trader now competes with the intermediary on her information.

3. Given the optimal strategy of the intermediary and the informed trader, we find  $\lambda$  by setting the risk-neutral market maker's expected profit equal to zero, i.e.

$$\begin{aligned} E[v|\omega] = \lambda\omega &\Leftrightarrow \frac{\text{cov}(\omega, v)}{\text{var}(\omega)} = \lambda\omega \\ \Leftrightarrow \lambda = \frac{\sqrt{3}\sigma_v}{2\sigma_u} > \frac{1}{2}\frac{\sigma_v}{\sigma_u} &\quad (\text{see equation (2.6)}) \end{aligned} \quad (2.11)$$

where  $\omega = ((\alpha + \beta)v + (1 + \gamma)u)$  is the aggregate order the market maker receives. Equation (2.11) shows that the price impact is increased in the presence of an intermediary and identifies two sources for the increased impact. First, the covariance in the numerator is increased due to more aggressive trading on the value  $v$ . We note, however, that the denominator is also increased to reflect the larger size of the order. Second, we find that the denominator is decreased due to less *net* “noise” trading as a result of the intermediary's strategy to trade opposite to the uninformed order.



## Appendix 2B: Additional Tables

**Table 2.B1: Macroeconomic Announcements**

This table describes scheduled macroeconomic announcements from 1994 to 1997. The data is from the International Money Market Services (MMS), except for the Housing Starts announcement dates which are from the Bureau of the Census ([www.census.gov](http://www.census.gov)) and FOMC announcement dates which are from Fleming and Piazzesi (2005). This table is modeled after Andersen et al. (2003, Table 1, p.43).

		Time (ET)	Number of Announcements				Total
			1994	1995	1996	1997	
<i>Quarterly</i>							
1	GDP Advance	08:30 a.m.	4	4	4	4	16
2	GDP Preliminary <sup>a</sup>	08:30 a.m.	4	3	4	4	15
3	GDP Final	08:30 a.m.	4	3	5	4	16
<i>Monthly</i>							
<i>Real Activity</i>							
4	Nonfarm Payroll Employment	08:30 a.m.	12	12	12	12	48
5	Retail Sales	08:30 a.m.	12	12	12	12	48
6	Industrial Production	09:15 a.m.	12	12	12	12	48
7	Capacity Utilization	09:15 a.m.	12	12	12	12	48
8	Personal Income <sup>b</sup>	08:30 a.m.	12	10	13	12	47
9	Consumer Credit Consumption	03:00 p.m.	0	0	7	12	19
10	Personal Consumption Expenditure <sup>c</sup>	08:30 a.m.	12	10	12	12	46
11	New Home Sales Investment	10:00 a.m.	12	11	13	12	48
12	Durable Goods Orders <sup>d</sup>	08:30 a.m.	12	11	13	12	48
13	Construction Spending	10:00 a.m.	12	12	12	12	48
14	Factory Orders	10:00 a.m.	12	12	12	12	48
15	Business Inventories <sup>e</sup>	10:00 a.m.	12	12	12	12	48
<i>Government Purchases</i>							
16	Government Budget <sup>f</sup>	02:00 p.m.	12	12	12	12	48
<i>Net Exports</i>							
17	Trade Balance Prices	08:30 a.m.	12	11	13	12	48
18	Producer Price Index	08:30 a.m.	12	12	12	12	48
19	Consumer Price Index	08:30 a.m.	12	12	12	12	48
<i>Forward Looking</i>							
20	Consumer Confidence Index	10:00 a.m.	12	12	12	12	48
21	NAPM Index	10:00 a.m.	12	12	12	12	48
22	Housing Starts	08:30 a.m.	12	12	12	12	48
23	Index of Leading Indicators	08:30 a.m.	12	11	13	12	48
<i>Six-Week</i>							
<i>FOMC</i>							
24	Target Federal Funds Rate <sup>g</sup>	02:15 p.m.	9	8	8	8	33
<i>Weekly</i>							
25	Initial Unemployment Claims	08:30 a.m.	52	52	52	53	209

<sup>a</sup>: Mar-96 missing; <sup>b</sup>: Jan-96 missing; <sup>c</sup>: Jan&Mar-96 missing; <sup>d</sup>: At 10:00 a.m. if same day GDP; Mar-96 missing; Jul-96 at 09:00; <sup>e</sup>: At 08:30 a.m. in 1997; <sup>f</sup>: Jan-96 missing; <sup>g</sup>: Around 02:15 p.m. no Expectations; 1994/02/04 was 11:15 announcement; 1994/04/18 was an unexpected announcement at 10:06; 1994/08/16 was at 01:17 and 1996/03/26 was at 11:39;

**Table 2.B2: Macroeconomic Announcement Surprise Coefficients**

This table reports the macroeconomic surprise coefficients from the profits regression of Table 2.3. In particular, we show estimation results of the following regression:

$$p_{t,h} - p_{t,h-1} = d_a(\alpha_a + \beta_a\omega_{t,h}) + d_n(\alpha_n + \beta_n\omega_{t,h}) + \sum_k \gamma_k I_{k,t} S_{k,t} + \varepsilon_{t,h}$$

where  $p_{t,h}$  is 100 times the log price of the 30Y treasury futures at day  $t$  and five minute interval  $h$ ,  $d_a$  ( $d_n$ ) is a dummy that is one on an announcement (nonannouncement) day, zero otherwise,  $\omega_{t,h}$  is the aggregate signed customer volume (CTI4, *Customer Flow*) divided by 1,000,  $S_{k,t}$  is the announcement surprise,  $I_{k,t}$  is a dummy that is one for the time interval immediately after the announcement, zero otherwise, and  $\varepsilon_{t,k}$  is the error term. For estimation, we use the Feasible Efficient GMM procedure with a Newey-West estimator (using three lags) for standard errors. We report the estimates of the announcement surprise coefficients estimated for the 8:30-8:35 interval.

Announcement Surprise Coefficients		
Announcement type	Surprise Coefficient	<i>t</i> -value
GDP Advance	-0.085*	-2.13
GDP Preliminary	-0.203	-1.68
GDP Final	0.036	1.91
Nonfarm Payroll Emp.	-0.469**	-3.06
Retail Sales	-0.111**	-2.95
Personal Income	-0.006	-0.27
Pers. Consumption Exp.	0.007	0.30
Dur. Goods Orders	-0.105**	-3.97
Business Inventories	-0.098*	-2.04
Net Exports	-0.006	-0.44
Producer Price Index	-0.175**	-4.41
Consumer Price Index	-0.121*	-2.54
Housing Starts	-0.105**	-5.32
Index of Leading Ind.	-0.018	-0.67
Init. Unemployment Cl.	0.043**	3.56

\*/\*\* indicates significance at the 95%/99% level.



# Chapter 3

## Are Market Makers Liquidity Suppliers?

This chapter is based on Van der Wel, Menkveld, and Sarkar (2008).

### 3.1 Introduction

Market makers play an important role in financial markets. They stand ready to buy and sell in order to accommodate the asynchronous arrival of sellers and buyers. Classic inventory models assume that they are risk-averse and therefore need compensation for carrying suboptimal inventory through time. In the process they earn the bid-ask spread to compensate for bearing inventory risk.<sup>1</sup> Empirically, studies have measured inventory control through the rate of inventory mean reversion and results are relatively poor, i.e. half-life of inventory typically is too long relative to what one might expect.<sup>2</sup>

Two likely explanations for slow inventory mean-reversion are (i) institutional features of a market that affect market maker behavior and (ii) active position-taking by market makers to speculate on private information. Panayides (2007) provides an example of the first and shows that the New York Stock Exchange specialist is sometimes forced to take on positions due to the Price Continuity rule. Madhavan and Smidt (1993) illustrate that a second reason for slow mean-reversion is speculation. They develop a model where the market maker actively manages an inventory to make it return to an optimal long-run

---

<sup>1</sup>The costs associated with the inventory risk is one of the three classic explanations for the bid-ask spread. The other two explanations are information asymmetry (see for example Kyle (1985) and Glosten and Milgrom (1985)) and order processing costs (such as in Roll (1984)). See O'Hara (1995) for an overview.

<sup>2</sup>For example, Hasbrouck and Sofianos (1993) show that it takes long to reduce an inventory position, sometimes up to two months. However, some of the inventory models' predictions are confirmed by data, most notably that market makers do manage inventory toward a target (see for example Manaster and Mann (1996) and Bjørnnes and Rime (2005)).

position, but, at the same time, they show that when he has access to private information he speculates by actively building a short-term position. This could explain the low level of mean-reversion if the econometrician ignores speculation.<sup>3</sup>

In this chapter, we explore whether (unconstrained) market makers actively speculate in the sense of Madhavan and Smidt (1993). We examine a large cross-section of 3,384 market makers active on the 30Y U.S. treasury futures trading pit on the Chicago Board of Trade. We find evidence of inventory control as the end of day inventory distribution is concentrated around zero. In addition, we find that indeed market makers actively take positions in the course of the day, i.e. they initiate trades that increase their inventory position. When we relate the extent of active position taking to proprietary trading profits we find a significant and positive correlation. This profitable position taking is consistent with active speculation by the market maker.

The current literature provides at best indirect evidence that market makers at times are speculators. Some studies find that the market maker initiates a significant share of his trades (Frino and Jarnecic (2000)) and interpret this as evidence of speculation. Initiation in itself, however, is not the same as speculation as market makers can initiate trades to actively manage their inventory position back to the long-term optimal level. A nice illustration of this phenomenon is the ‘hot-potato’ trading model of Lyons (1997). Locke and Sarajoti (2004) find market maker inventory can be split into a desired and undesired position (from a long-term rational expectations point of view), and that market makers are most aggressive to offset their undesired position. Manaster and Mann (1996) and Bjønnes and Rime (2005) document strong inventory control but, surprisingly, do not find the expected price effects of inventory.<sup>4</sup> They conclude that their results are consistent with active speculative position taking by floor traders. Anand and Subrahmanyam (2007) show that intermediaries, a subset of the market makers we study, have information orthogonal to their clients and account for greater price discovery. And, contrary to all these studies, Chakravary and Li (2003) study market making but do not find evidence in favor of speculation. We add to all of the aforementioned papers, as we are the first to provide direct evidence of active position taking by market makers. Moreover, when we relate this position taking to trading profits we find a positive and significant relation.

To study the liquidity demand and profitable position taking of the market maker further we disentangle our main result in two dimensions. First, as we find that market makers can be informed speculators it is interesting to distinguish days with a high-

---

<sup>3</sup>One outcome of their model is an equation that explicitly explains today’s inventory as the sum of two components: (i) a fraction of yesterday’s inventory consistent with inventory management and (ii) an active position on private information. We discuss the equation further in the literature section.

<sup>4</sup>Demsetz (1968) and Stoll (1978) show that it could be optimal for market makers to not quote symmetrically around the efficient price, but deviate in order to create incentives for liquidity demanders to trade in the direction that brings the market maker inventory back to zero.

information environment and days with a low-information environment. If market makers are able to trade on information we expect them to be more active in taking positions on days at which there is more information coming to the market, and that their trading profits from these positions are higher. To identify high- and low-information environments we use the days on which there are the scheduled releases of macroeconomic news. This is a contribution in itself: though many studies find that these scheduled releases of news significantly affect returns, volatility, volume and information asymmetry<sup>5</sup> the studies that look at liquidity supply<sup>6</sup> do not take these announcements into account.

We find empirically that the floor traders prefer a flat position before the announcement, and quickly build up a position after the news is released. In anticipation of the scheduled news release the inventory distribution is highly concentrated around zero. After the announcement market makers quickly build up their position, both through actively initiating trades and passively participating in trades that increase their inventory positions. The inventory positions on announcement days are larger than on nonannouncement days, confirming our expectation. If we relate the extent to which they participate (actively and passively) in trades that increase their inventory position to their profits from trading we find a significant positive correlation. This is consistent with the market makers demanding liquidity on days with high-information environment, and speculating right after the news announcement.

The second dimension with which we study the main result is the heterogeneity in information sets of the various market makers. Of the total group of 3,384 market makers on the average day there are 521 traders active, of which 155 traders trade both for own account and on behalf of customers. This last group of traders is commonly referred to as dual traders, or simply duals. Recent studies find that order flow coming from the customers of dual traders, the end-users in the economy, provides information orthogonal to that from the macroeconomic news (see Chapter 2), and may even predict macroeconomic variables (Evans and Lyons (2005, 2007)). Similar to looking at days with high- and low-information environment we expect the traders with the largest information set to be most active in building up positions, and that they obtain the highest profits from this position taking. Studying subgroups in the population of market makers is also consistent with the results of Manaster and Mann (1996, p.973), who point out that “market makers are not merely passive order fillers, ..., but are active profit-seeking individuals

---

<sup>5</sup>See Ederington and Lee (1993, 1995), Fleming and Remolona (1997, 1999), Balduzzi, Elton, and Green (2001) and Andersen, Bollerslev, Diebold, and Vega (2003, 2007) for evidence on the effects of macroeconomic announcements on returns, volatility and trading volume. Chapter 2 documents higher information asymmetry after macroeconomic announcements, consistent with Green (2004) and Pasquariello and Vega (2007).

<sup>6</sup>Such as Manaster and Mann (1996) and Chakravary and Li (2003) for the futures market, see Section 3.2 for a full overview.

with heterogeneous levels of information and/or trading skill". We are the first to explicitly pick up heterogeneous subgroups in the context of liquidity supply and demand by comparing these dual traders with traders that only trade for their own account ('local' traders).<sup>7</sup>

We find that the market makers with the additional information of observing order flow from customers are more active in building up inventory positions after macroeconomic announcements. Moreover, for this subgroup of dual traders the relation between the amount of trades that increase their inventory position and trading profits is strongest. This is consistent with the dual traders using the information they obtain from bringing customer orders to the market and speculating based on this.

Our dataset of the 30Y U.S. treasury futures in 1994-1997 has three advantages. First, we can identify whether trades are proprietary or on behalf of customers. After signing the trades as buyer- or seller-initiated this allows us to study not only liquidity supply, but also liquidity demand of the market maker. Second, of the work that focuses solely on the liquidity supplying role of the market maker the majority looks at the New York Stock Exchange Specialist,<sup>8</sup> who, as was pointed out earlier, is not a pure market maker as he is sometimes restricted by his obligation to smooth transaction prices. Therefore it is more natural to study the market maker in settings where he does not have such an obligation, as in treasury futures markets. Third, an advantage of our dataset over the existing treasury futures market liquidity supply studies is that the maturity we look at has the largest share of trading at one market. Manaster and Mann (1996) and Chakravary and Li (2003) study liquidity supply on the market for the less liquid 13 week bill, on which trading is split between the spot and futures market. When this is the case one has to account for hedging across markets.<sup>9</sup>

Though in recent years on many markets other market participants supply liquidity through limit orders, the majority of markets still rely on a subset of traders that always stand ready to trade (of the 51 exchanges that Jain (2005) studies 27 have a dealer emphasis or hybrid system). The trend in financial markets is toward electronic trading platforms, but also here there are market makers and intermediaries active like the ones we study in this chapter. For example, Hendershott, Jones, and Menkveld (2007) study

---

<sup>7</sup>Others have compared profits of dual traders to profits of local traders (see, e.g., Fishman and Longstaff (1992)), but not inventory management of these two groups of traders.

<sup>8</sup>As in Madhavan and Smidt (1991), Hasbrouck and Sofianos (1993), Madhavan and Sofianos (1998) and Hendershott and Seasholes (2007), see Section 3.2 for a full overview.

<sup>9</sup>For the 13 week bill about 55% of volume trades on the futures market. For the 30 year bond almost all trading takes place on the futures market, about 95%. These calculations are based on Fleming and Sarkar (1999). For spot the on-the-run security is taken, for futures the nearby contract. The sample size of both studies illustrate the lower liquidity of the 13 week contract: Manaster and Mann (1996) look at 584 trader days, Chakravary and Li (2003) study only 6 traders for the treasury futures market (though both studies include more data, but from other futures contracts).

algorithmic trading and emphasize the role of algorithmic traders as liquidity suppliers. In addition, recently markets are interested in the introduction of Designated Market Makers, who commit to providing a minimum level of liquidity.<sup>10</sup>

In the implementation we come across the challenge of signing transactions in the absence of quotes. Hasbrouck (2004) suggests implementing a signing method based on the Roll (1984) model that relies on a Bayesian inference technique, the Gibbs Sampler. As this is a simulation based method it might be computationally burdensome, and is too time-consuming for our dataset of over 42 million trades. The second main contribution of this chapter is that we develop an innovative method for determining whether the buying or selling party initiated a trade. The method we propose follows from the same principles as Hasbrouck's (2004) method. However, as our model relies on methods of time-series models in state space form we greatly reduce the computational burden. Applying Hasbrouck's method to our dataset would require over 40 hours for one model variation to be estimated, whereas the method we suggest requires a little over 110 minutes.<sup>11</sup>

Overall our results imply that the market maker is not an uninformed liquidity supplier. He initiates a significant part of his trades, and sometimes actively builds up a speculative position. This is particularly true when looking at the high-information environment created by the scheduled releases of macroeconomic news and for traders with a large information set. The market making behavior is consistent with the Madhavan and Smidt (1993) model in which the market maker is both a dealer and a speculator. Our results emphasize the need for theoretical models that take the informativeness of the market maker into account, such as the model recently put forward by Boulatov and George (2007).

The rest of the chapter is build up as follows. Section 3.2 discusses related literature. In Section 3.3 we discuss our dataset, provide some institutional background, and show summary statistics. Section 3.4 details the method we use to sign the trades in our dataset. In Section 3.5 we present our empirical findings. Section 3.6 concludes.

---

<sup>10</sup>See Bessembinder, Hao, and Lemmon (2007), Venkataraman and Waisburd (2007), Anand, Tanggaard, and Weaver (2008) and Menkveld and Wang (2008).

<sup>11</sup>The number of times each of the methods needs to run over all the observations causes the difference in required calculation time. Hasbrouck (2004) suggests 10,000 swoops over the data, while our likelihood-based method requires on average 10 maximum likelihood iterations to estimate the two parameters in which the likelihood is calculated about 6 times. See Section 3.4.2 for more details on the differences in estimation procedures and calculation times.



## 3.2 Related Literature

### 3.2.1 Market Making and Inventory

Our study of the market maker relates to several strands of literature. First, it relates to the literature that assumes the market maker is an uninformed liquidity supplier. For example, in Stoll (1978) the market maker adjusts his quotes depending on his inventory position to induce mean-reversion in inventory and earns the spread. In the adverse selection models of Kyle (1985) and Glosten and Milgrom (1985) the market maker sets prices such that in expectation he has no profits and absorbs the net order flow. In both types of models the market maker is uninformed and a liquidity supplier.

Madhavan and Smidt (1993) are the first to provide an exception to the above setting. They introduce a model in which the market maker is both a dealer and a speculator. Equation (6) in their paper gives the optimal trade quantity of the market maker, and illustrates this dual nature:

$$I_{t+1} - I_t = \beta(I_t - I^d) - \frac{1 + \beta}{2}x_t,$$

with  $I_t$  the market maker's inventory position at time  $t$ ,  $-1 < \beta < 0$  a parameter measuring the speed of inventory adjustment,  $I^d$  the long-term desired inventory position and  $x_t$  representing the short-horizon investment strategy. Thus in the Madhavan and Smidt (1993) model the optimal trade quantity of the market maker consists of two components: the deviation of his inventory from the long-run desired level and a short-run speculative strategy.

In Boulatov and George (2007) liquidity suppliers can also be informed. Specifically, informed traders may choose whether to act as liquidity demanders or suppliers. In case of fully anonymous liquidity provision the informed act exclusively as liquidity suppliers. If there is less than full anonymity some informed traders choose to provide liquidity while others demand liquidity. Hendershott, Jones, and Menkveld (2007) provide empirical results consistent with this model, as they find large informativeness of quotes. Our results offer empirical guidance on whether market makers can indeed be informed.

Second, we relate to empirical studies of liquidity supply and inventory management. The great majority of these papers look at the Specialist on the New York Stock Exchange (Hasbrouck (1988), Madhavan and Smidt (1991), Hasbrouck and Sofianos (1993), Madhavan and Sofianos (1998), Panayides (2007) and Hendershott and Seasholes (2007)). In addition researchers look at the U.S. treasury market (Manaster and Mann (1996) and Chakravary and Li (2003)), the exchange rate market (Lyons (1995), Bjønnes and Rime (2005) and Cao, Evans, and Lyons (2006)), the London Stock Exchange (Hansch, Naik,

and Viswanathan (1998), Reiss and Werner (1998) and Naik and Yadav (2003)) and option data (Garleanu, Pedersen, and Poteshman (2005)).

We differ from these studies in a number of issues. First, we do not take the liquidity supply role of the market maker as given, but also consider the liquidity demand role of the market maker and examine whether there indeed is a liquidity supply role. Second, we look at the treasury market in which there is a large cross-section of market makers active with large heterogeneity. These market makers are free of obligations, whereas the NYSE specialist is sometimes restricted by his other obligations such as smoothing transaction prices (Panayides (2007)). Compared to the current treasury market studies we distinguish ourselves also by looking at a maturity at which trading is concentrated on one exchange. In addition our market is more active, such that we have a larger sample of floor traders to examine (see Footnote 9 for more on these points).

Manaster and Mann (1996) and Bjønnes and Rime (2005) document strong evidence of inventory control by market makers. Surprisingly however, they do not find price effects of the inventory positions. They interpret this as being consistent with active position taking by floor traders. We add to this, and show direct evidence of active and passive position taking by market makers. Moreover, we find floor traders that are most active in building up a position earn higher profits from trading.

### 3.2.2 Signing Trades in the Absence of Quotes

We also relate to the literature on classifying trades according to whether they are initiated by the buying or selling party (the so-called ‘signing’ of trades). For markets with explicit quotes a trade is qualified as being initiated by the buying party if it takes place closer to the ask than the bid (Hasbrouck (1988), Lee and Ready (1991) and Ellis, Michaely, and O’Hara (2000)).

For markets without quotes identifying whether the buying or selling party initiated the trade is more challenging as it is difficult to find a good ‘reference’ for the observed transaction prices. A tick test can be used, where a trade is labeled as being initiated by the buying party if it is an uptick (i.e. if the transaction price is larger than the previous price). However, this method has the disadvantage that a trade can be incorrectly labeled as buyer-initiated simply due to an orthogonal price innovation. Alternatively, Rosenberg and Traub (2007) suggest to use the quotes of a parallel market for the same asset as a reference: to sign futures market trades they use the quote from the forward market. Unfortunately, this is not applicable in general as there needs to be such a parallel market.

Hasbrouck (2004) suggests using a Bayesian methodology (the Gibbs sampler) to explicitly model the price innovation and trade sign. Unlike the aforementioned methods this grounds in economic theory: it is based on the Roll (1984) model. We follow his

approach, but suggest a much quicker likelihood based algorithm to sign futures market trades. In addition we develop a method that deals with high-frequency datasets in which it can occur that there are multiple trades occurring at the same second at the same price.

### 3.3 Data and Institutional Background

To study the market maker we analyze trading in the 30Y U.S. treasury futures pit. This instrument trades on the Chicago Board of Trade (CBOT). On the trading floor (the ‘pit’) traders (‘floor traders’) gather between 8:20 a.m. and 3:00 p.m. Eastern Time. Trading takes place through the open outcry method: prices are negotiated by shouting out orders and indicating direction and quantity using hand signals.

We look at the 30Y maturity (instead of e.g. the 5Y) as this is the maturity where volume is most concentrated on one exchange. For the 30Y maturity about 95% takes place on the futures market and 5% on the spot, while for the 5Y bond this is 24% and 76% respectively (see Fleming and Sarkar (1999)). In addition, in our sample period of 1994-1997 electronic trading is still limited. Thus, by looking at the 30Y treasury futures we have a setting where we observe almost all trading, and minimize the risk of missing offsetting trades in the other market or electronically.

At each moment in time multiple treasury bond futures contracts with different expiry months are traded. We focus on the most nearby as this is the most liquid of these (see Fleming and Sarkar (1999)) and makes it a very close substitute for the underlying spot instrument. As Ederington and Lee (1993, p.1164) point out, this makes our results generalizable to the spot market.

Our dataset records all trades taking place on the futures pit. For each trade are recorded: the time of the trade; a buy/sell indicator; trade quantity (in contracts); trade price; a floor trader identifier and a customer type indicator (CTI). Floor traders have to report their trades in 15 minute brackets. A timing algorithm (the Computerized Trade Reconstruction) is used to time the reported trades to their nearest second. Though this may be noisy we believe this timing is fairly accurate.<sup>12</sup> It is used by the Commodity Futures Trading Commission (CFTC) for regulation purposes, and is used in e.g. the studies of Fishman and Longstaff (1992) and Manaster and Mann (1996).

The CTI indicates for each trade whether it is a trade for the own account of the floor trader (a ‘proprietary’ trade), or on behalf of another party. In particular, we have the

---

<sup>12</sup>In addition, we apply several data filters. First, we focus on ‘regular’ trades: we remove trades that are e.g. indicated to be spread trades. Second, we remove trades that show an unusual transaction pattern. Specifically, if a transaction return of more than 0.25% is followed by a return in the opposite direction also larger than 0.25% we expect these trades to suffer from serious timing error and eliminate it. In total we remove 1.48% of all trades with these two filters.

following four codes: CTI1: proprietary trade; CTI2: trade for clearing member's house account; CTI3: trade for another member present at the floor; CTI 4: a trade for (off-exchange) customers. Consistent with earlier futures market studies (such as Fishman and Longstaff (1992), Manaster and Mann (1996) and Chakravary and Li (2003)) we restrict attention to CTI1 and CTI4 trades as they represent most trading volume. Since both parties report a trade the trades are double counted. For example, a trade between two market makers each trading for own account appears twice in our dataset with both times the same quantity, price and CTI but different buy/sell indicator and floor trader identifier.

Following the futures market literature (see for example Fishman and Longstaff (1992), Chang, Locke, and Mann (1994), Manaster and Mann (1996) and Chakravary and Li (2003)) we use the CTI codes to identify three groups of floor traders in our data. On a daily basis we identify traders that only trade on behalf of customers (we label these as 'brokers'), traders that only trade for own account ('locals', or local traders) and traders that do both ('duals', or dual traders).<sup>13</sup> Table 3.1 provides some summary statistics for these groups. Important to note is that of these three groups the locals and duals trade for own account, and are the market makers in this setting.

On an average day there are 521 traders active, each generating an average volume of 807 contracts for own account and 793 contracts on behalf of customers. This illustrates the enormous activity of this market: on an average day 357 thousand contracts are traded for own account, and 184 thousand contracts are traded on behalf of customers. Of the 521 traders active on an average day 288 are local traders, 155 dual traders and 78 brokers. Thus, on an average day in our sample 443 floor traders provide market making services.

We combine our above treasury futures dataset with a dataset on macroeconomic announcements from the International Money Market Services (MMS). We consider a broad set of 25 macro announcements such as the PPI, CPI and Nonfarm Payroll Employment figures (which I look at in Chapter 2 and have previously been studied by, e.g., Green (2004) and Pasquariello and Vega (2007)) that occur throughout the trading day at fixed times, such as at 8:30 a.m. ET and 10:00 a.m. ET. For each macro announcement an expectation of market participants is recorded, together with the first released (i.e., not revised) figure.

Consistent with the aforementioned studies, we focus on the 8:30 announcements as this is where the most significant announcements are. Using the announcement data we

---

<sup>13</sup>Following the literature we allow for a 2% error margin for this classification (see for example Fishman and Longstaff (1992) and Chang, Locke, and Mann (1994)). That is, if daily volume for a trader on a day consists of more than 98% (less than 2%) proprietary volume we label him a local (broker), otherwise a dual.

**Table 3.1: Summary Statistics**

This table shows the trading activity for locals, duals and brokers on the market for the 30Y treasury futures in 1994-1997. We classify traders at the daily basis, and label a trader to be a local (broker) if more than 98% (less than 2%) of his trades are for own account, otherwise he is a dual. For these three groups we show the average number of days a trader is active, the average number of traders active on a day, the total number of trading days and the average daily volume per trader. In addition we show these for the three groups of traders combined (*All Traders*). The column *Full Sample* shows the total number of days and number of traders observed in our sample.

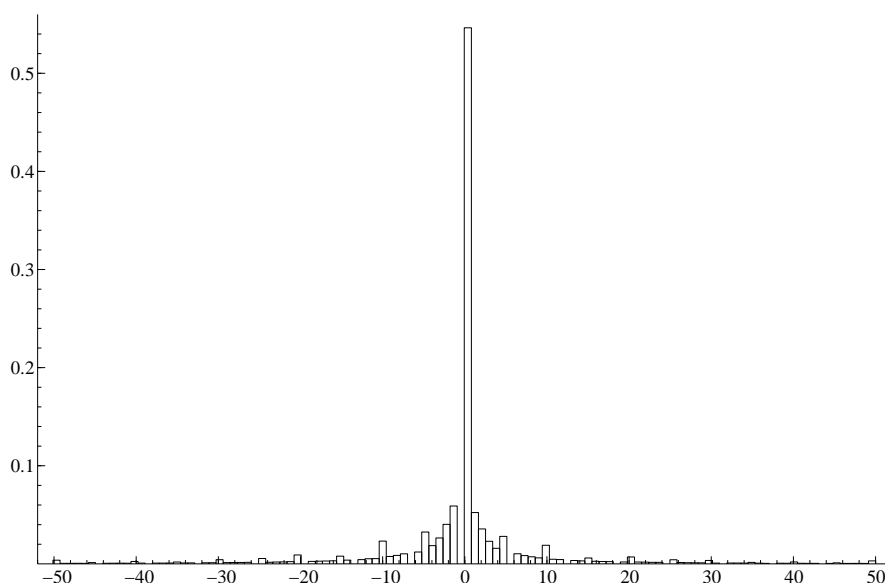
Summary Statistics - Trader Activity					
	Full	All			
	Sample	Traders	Locals	Duals	Brokers
Avg #days a trader is active					
all days	1,005	155	173	130	27
ann days	377	60	67	50	11
nonann days	350	52	58	44	9
Avg #traders active per day					
all days	3,384	521	288	155	78
ann days		535	296	157	82
nonann days		501	276	150	74
Total number of trading days					
all days		523,540	289,354	155,381	78,805
ann days		201,571	111,645	59,199	30,727
nonann days		175,349	96,674	52,664	26,011
Average daily volume per trader					
For own account					
all days		807	1,070	318	
ann days		903	1,198	345	
nonann days		682	897	288	
For customers					
all days		793		921	541
ann days		894		1,048	596
nonann days		667		766	467

split our sample in two groups. We look at days at which there is one or more 8:30 announcement, and days at which there are no announcements at 8:30. To make sure the 8:30 announcement is driving the results we are careful to take out days with other announcements that take place in the morning (that is, at 9:15 and 10:00).

Table 3.1 also provides summary statistics for these announcement and nonannouncement days. Of the total 1,005 days in our sample we find there are 377 announcement days, and 350 nonannouncement days. We see a clear increase in trading activity both in average volume traded and number of traders on announcement days compared to nonannouncement days. For example, on average on announcement days there are 535 traders active, while on nonannouncement days there are 501 traders, 6.4% less.

### Figure 3.1: End of Day Inventory

This figure shows the end of day inventory for floor traders active in the 30Y treasury futures in 1994-1997. For each day the end of day inventory position is calculated only for traders that were active on that day, assuming a zero inventory position at the beginning of the day. The figure shows the histogram of the end of day inventories.



For each trader we have a record of all his trades in the 30Y treasury futures, plus the direction of each of these trades. We use this to obtain the inventory for each trader. Consistent with the previous literature (see for example Manaster and Mann (1996)) we do this under the assumption that floor traders close out the day with zero inventory.<sup>14</sup>

<sup>14</sup>There are some limitations to this way of calculating inventories. As we focus on CTI1 and CTI4 trades we miss possible CTI2 and CTI3 trades of market makers. In addition we only have a record of the trades in the 30Y treasury futures, and not in other markets. But similar to Manaster and Mann

In Figure 3.1 we plot the end of day inventory that is obtained using this assumption. From the figure it is clear that end of day inventory is indeed centered around zero. The most common end of day inventory position is a flat position. Of the nonzero positions most are below 15 contracts in absolute terms, which is small compared to the average market maker's trade size of 807 contracts. This suggests that the assumption that is used to construct the inventory series is a reasonable one.

## 3.4 Signing Trades in the Absence of Quotes

### 3.4.1 Methodology

To study the behavior of market makers it is necessary to identify whether the trades in which a market maker is involved are initiated by his counterparty or by himself. This amounts to determining for each trade whether it is initiated by the buying or selling party. For markets with explicit bid and ask quotes algorithms for this challenging task are available, an often employed technique that relates the transaction price to the average of the bid and ask quote (the 'midquote') was put forward by Lee and Ready (1991). For markets without quotes the identification is an even more challenging task. Whereas in the former case the observations consist of transaction prices and both bid and ask quotes, for the latter only the transaction prices are observed.

Hasbrouck (2004) proposes a new Bayesian methodology to deal with the challenge of estimating the unobserved sign of the trades from the observed transaction prices. The methodology is based on the Roll (1984) model of the bid-ask spread.<sup>15</sup> In this model the logarithm of the unobserved efficient price  $m_t$  evolves as a random walk:

$$m_t = m_{t-1} + u_t, \quad u_t \sim N(0, \sigma_u^2). \quad (3.1)$$

The actually observed log transaction prices  $p_t$  are either above or below this unobserved efficient price, depending on whether the trade is initiated by the buyer or the seller of the transaction. If we let  $q_t \in \{-1, +1\}$  denote the direction of a trade, with  $+1$  a buyer-initiated trade and  $-1$  a seller-initiated trade, and  $c$  the transaction costs we can write

---

(1996), we believe the current method provides the most accurate estimate of inventories.

<sup>15</sup>In addition, Hasbrouck (2004) suggests several extensions to the Roll (1984) model that can be estimated in the Bayesian framework he proposes. As our main purpose is the signing of trades we choose a standard setting and therefore remain in the set-up of the Roll (1984) model. Note that in addition to this standard setting, in Footnote 19 we develop an alternative method to sign futures trades that allows for more flexibility.

the observed log transaction price as:

$$p_t = \begin{cases} m_t + c, & \text{if } q_t = +1, \\ m_t - c, & \text{if } q_t = -1, \end{cases} \quad (3.2)$$

or simply  $p_t = m_t + cq_t$ . We note that  $q_t$  is not directly observed.

Hasbrouck (2004) suggests a Markov Chain Monte Carlo (MCMC) methodology (the Gibbs Sampler), in which iteratively draws from the parameters  $c$  and  $\sigma_u^2$  are obtained, together with draws for the unobserved time series of signs  $q_t$ ,  $t = 1, \dots, n$ . For a large number of draws the simulated distribution will be equal to the desired joint posterior distribution of the parameters and sign time series, conditional on the observed prices. An obvious disadvantage of these simulation based techniques is that they are computationally expensive and require a lot of simulations to obtain convergence. As we have over 42.5 million observations this MCMC technique requires long computation time.

Instead of the above Bayesian methodology we propose to estimate the Roll (1984) model parameters  $c$  and  $\sigma_u^2$  and the series of trade initiating signs  $q_t$  in a State Space Framework (SSF). This class of time-series models builds on the idea that an observed series can be explained by several unobserved components.<sup>16</sup> If we write the Roll (1984) model in this framework we obtain:

$$\begin{aligned} p_t &= m_t + cq_t, & q_t &\in \{-1, +1\}, \\ m_t &= m_{t-1} + u_t, & u_t &\sim N(0, \sigma_u^2). \end{aligned} \quad (3.3)$$

As the  $q_t$  are not Gaussian but binary distributed (assumed to be initiated by the buying or selling party with equal probability<sup>17</sup>) this is not a standard linear Gaussian state space model. However, since  $q_t$  can only take on two values this model can be seen as a special case of state space models with regime switching. Kim and Nelson (1999) discuss the implementation of regime switching models in the state space framework, and show that it is a combination of the Kalman filter for SSF models and the Hamilton (1989) filter for regime switching models. By implementing the recursions of this algorithm we obtain a likelihood value, which we can maximize using standard optimization techniques. See Appendix 3A for details on the estimation procedure.

---

<sup>16</sup>See Durbin and Koopman (2001) for an introduction. Some recent financial market studies that use state space techniques are Driessen (2005) and Menkveld, Koopman, and Lucas (2007); Glosten and Harris (1988), Harris (1990) and Hasbrouck (1999) employ non-Gaussian state space methods.

<sup>17</sup>This is an assumption that can easily be relaxed in this framework, which we see as a major contribution of our methodology. Whereas the extensions that Hasbrouck (2004) proposes deal with the price impact of trades, trade clustering and price discreteness, we are also able to straightforwardly extend the model in terms of autocorrelation in sign.



### 3.4.2 Simulation Study and Results

There are a few additional issues when applying our State Space Form Regime Switching (*SSFRegSw*) signing methodology to the data introduced in Section 3.3. As our dataset deals with a very actively traded instrument it occurs that there are multiple trades at the same second.<sup>18</sup> As in the Roll (1984) model the transaction cost and efficient price innovation and thus the sign does not depend on quantity we summarize the information of all trades in this second to one single observation. In addition, as was pointed out in Section 3.3, the time stamps of the trades are obtained using the Computerized Trade Reconciliation algorithm and are therefore noisy to some extent. Though this is not an issue addressed in the Hasbrouck (2004) methodology, for robustness we implement an alternative signing algorithm also based on the state space model, in which we aggregate all trades within each minute.<sup>19</sup> Our results remain unchanged.

Panel A of Table 3.2 provides results from a simulation study of the accuracy and speed of the various signing algorithms. For a number of replications (100 in this set-up) we generate a fixed number of observations (chosen to be 50 here) with Roll (1984) as the Data Generating Process (DGP). For each of these replications we estimate the parameters of the Roll model and the sign (buyer- or seller-initiated) of each trade. We compare the following methods: the method of moments,<sup>20</sup> the State Space Form Regime Switching (*SSFRegSw*) methodology from the previous section, the SSF Approximation (*SSFApprox*) method from Footnote 19, Hasbrouck's (2004) MCMC (*H-MCMC*) method and a tick test (in which a trade is considered to be buyer-initiated if it is an uptick).

We find that the method of moments, SSF Regime Switching and Hasbrouck's method all provide good estimates of the Roll model parameters. Moreover, the results for the SSF Regime Switching and Hasbrouck's method are very similar. For some parameter

---

<sup>18</sup>It can even occur that in the futures pit there are trades at different prices in the same second. On average this occurs 28 seconds per day (at about 1.1% of all seconds at which trading takes place). Though in the *SSFRegSw* signing framework this problem is ignored, we address this in the robustness signing algorithm of Footnote 19.

<sup>19</sup>We approximate the Roll (1984) model of equation (3.3) with a linear Gaussian state space model:

$$\begin{aligned} p_t &= m_t + v_t, & v_t &\sim N(0, \sigma_v^2), \\ m_t &= m_{t-1} + u_t, & u_t &\sim N(0, \sigma_u^2), \end{aligned}$$

such that we have  $\mathbb{E}[v_t] = \mathbb{E}[u_t] = 0$  and  $\mathbb{V}[v_t] = \mathbb{V}[u_t] = 2c$ . As we now do not need the additional calculations of the Hamilton (1989) filter we can straightforwardly implement a multivariate version of this model. This allows us to aggregate the trades within a certain interval (10 seconds for example, or 1 minute) by creating a multivariate price vector for each time and keeping  $m_t$  as a scalar. That this creates a variable number of observations is no problem for models in SSF, as this class of models is particularly well-suited for models with missings. In addition we can also use this approach to deal with the situation that in some seconds there are trades occurring at different prices.

<sup>20</sup>This methodology works as follows. For the difference of the price process in (3.3) we have  $\mathbb{V}ar(\Delta p_t) = 2c^2 + \sigma_u^2$  and  $\mathbb{C}ov(\Delta p_{t-1}, \Delta p_t) = -c^2$ . By matching these moments with the estimates from the data we obtain estimates for  $c$  and  $\sigma_u^2$ . Note that we do not obtain estimates for  $m_t$  and  $q_t$  in this case.

**Table 3.2: Signing Futures Market Trades - Simulation Study and Results**

This table shows the results for signing the futures market trades on the market for the 30Y treasury futures in 1994-1997. We compare the output from the State Space Form Regime Switching (*SSFRegSw*) signing algorithm we propose to the State Space Form approximation method (*SSFApprox*), Hasbrouck's (2004) MCMC (*H-MCMC*) method and a tick test (in which a trade is labeled as being initiated by the buying party if it is an uptick). In Panel A we show the output from a simulation study of these signing methods. The parameters we use for our data generation process (DGP) are chosen such that they are close to values actually observed in the data. For the five signing methods we report the mean and standard deviation (*St Dev*) of the half-spread  $c$  and the efficient price variance (*Eff Price Var*), the percentage of trades that are signed correctly, the root mean squared error (*RMSE*, x1,000,000) of the smoothed efficient price versus the true value and the time needed to run the algorithm (in seconds). For the simulation 50 observations and 100 replications are used. In Panel B we report results from running various algorithms on 10 days of our dataset. We look at the *SSFRegSw* method, *SSFApprox* with aggregation of both 10 and 60 seconds (see Footnote 19) and the *H-MCMC* method. We show the mean and standard deviation (*St Dev*) of the half-spread  $c$  and the efficient price variance (*Eff Price Var*), the percentage of trades that are labeled the same as the *SSFRegSw* method and the average time needed to obtain these results for one day (in seconds).

Panel A: Signing Futures Market Trades - Simulation Study							
	Half-Spread		Eff Price Var		% Sign Correct	RMSE of $m_t$	Calc Time
	(c; x1,000)		( $\sigma_u^2$ ; x1,000,000)				
	Mean	St Dev	Mean	St Dev			
Parameters DGP	0.200		0.010				
Method of Moments	0.199	0.034	0.009	0.019			0.0
SSF Regime Switching	0.199	0.011	0.010	0.002	95.7	0.290	0.4
SSF Approximation	0.089	0.036	0.041	0.007	95.1	0.550	0.0
Hasbrouck MCMC	0.153	0.046	0.036	0.019	95.6	0.422	3.9
Tick Test					73.6		0.0

Panel B: Signing Futures Market Trades - Results for 10 Days						
	Half-Spread		Eff Price Var		% Sign Same as RegSw	Calc Time
	(c; x1,000)		( $\sigma_u^2$ ; x1,000,000)			
	Mean	St Dev	Mean	St Dev		
SSF Regime Switching	0.158	0.002	0.017	0.004		6.7
SSF Approx (10s)	0.091	0.030	0.042	0.005	89.1	1.4
SSF Approx (60s)	0.192	0.071	0.089	0.035	81.9	0.3
Hasbrouck MCMC	0.157	0.003	0.019	0.005	86.7	143.3

values and number of observations our SSFRegSw method performs slightly better, for other parameter values and number of observations the H-MCMC method performs a bit better.<sup>21</sup> However, the SSF Approximating method performs very poorly to estimate the model parameters.

In terms of estimating the sign of the trades we see both the SSF Regime Switching method and Hasbrouck's MCMC method perform very well, with an accuracy of over 95% and greatly outperforming the tick test. The MCMC and SSFRegSw methods provide the most accurate results, signing at least another extra 0.5% of trades correct compared to the competing methods. Interestingly, while the SSF Approximating method performs poor in obtaining the model parameters, it seems to work very well for signing trades.

In terms of time there is a clear difference between the SSF methods and the MCMC method. Hasbrouck's MCMC method is more than 10 times slower than the SSF Regime Switching method. As the MCMC method relies on simulation methods more loops over the data are needed. For example, taking 10,000 swoops (as Hasbrouck recommends) requires  $10,000 * n$  calculations, times the number of calculations at every  $t = 1, \dots, n$ . The Maximum Likelihood (ML) calculations based on the SSF Regime Switching method require less. On average 10 iterations in the ML procedure are needed to obtain convergence, in which on average 6 times the likelihood has to be calculated. Therefore on average the ML method requires about  $10 * 6 * n = 60 * n$  calculations times the number of calculations at every  $t = 1, \dots, n$ . Even without considering the number of calculations at every  $t = 1, \dots, n$  for both methods it is clear that the SSFRegSw method is much quicker than the MCMC method. The SSF Approximation method provides the quickest results.<sup>22</sup>

Overall there is a case to be made for both of the SSF signing methods and the MCMC method. As our dataset contains many trades we prefer to take the quickest of these methods, and employ the SSF Regime Switching method. In addition we use the SSF Approximation method for robustness.

Figure 3.2 illustrates how we apply the SSF Regime Switching signing methodology to the data. In the top figure we plot the raw data: the observed transaction prices. Using these observations we obtain a smoothed efficient price series (also plotted in the top figure) and a smoothed probability of the trade being initiated by the buying party.

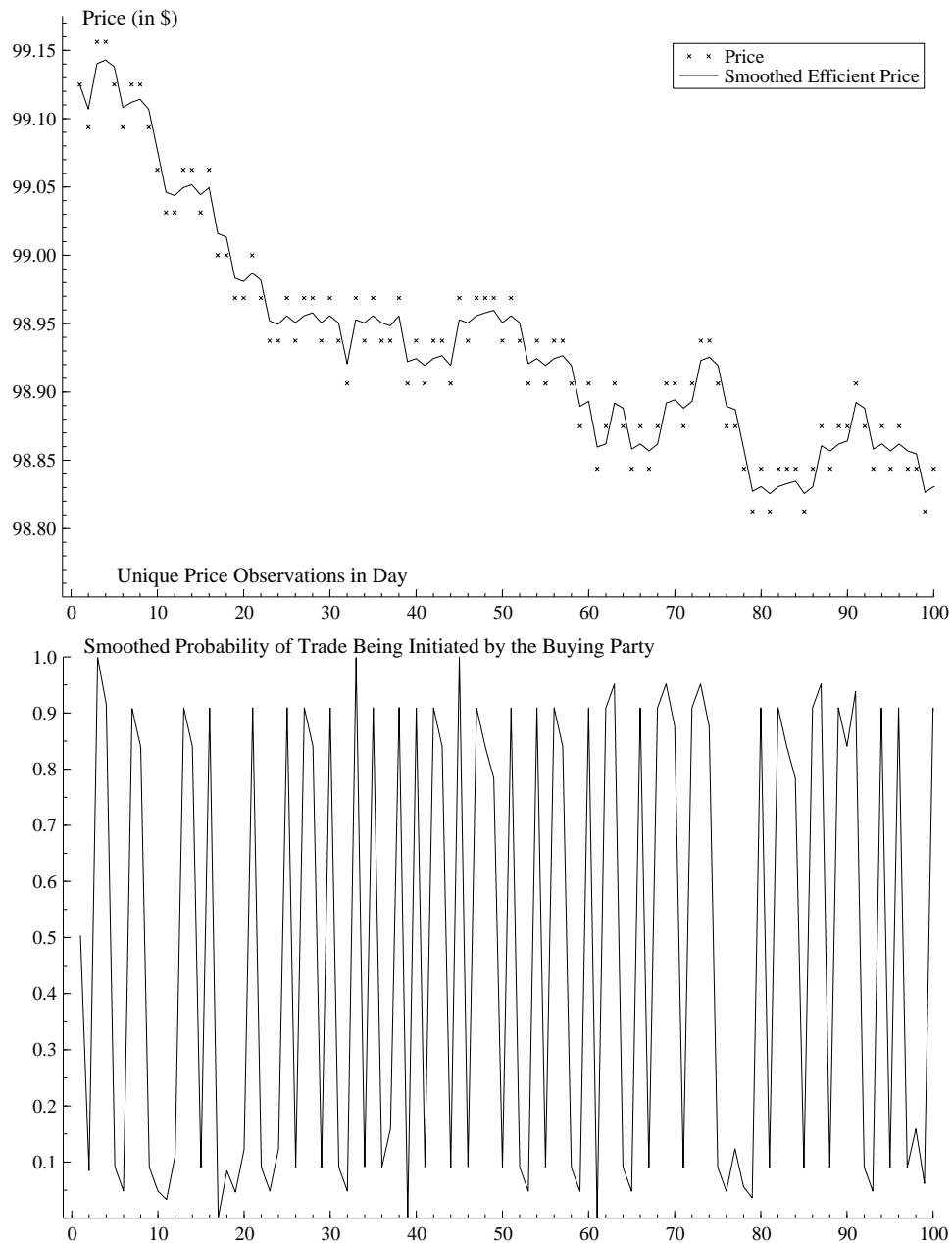
---

<sup>21</sup>From the results reported in the table it may seem that the Hasbrouck (2004) MCMC method is more biased and less efficient than the SSF Regime Switching method. However, this is due to the chosen number of swoops and burn-in of the MCMC method. With a greater burn-in results similar to the SSFRegSw can be obtained with the H-MCMC method, though this will add to the calculation time.

<sup>22</sup>This is however partially due to the underlying code. The Hasbrouck (2004) MCMC and State Space Form Regime Switching methods are straightforwardly implemented in Ox (see Doornik (1998)), while the SSF Approximation method uses the functions from the SsfPack (see Koopman, Shephard, and Doornik (1999)) which are programmed in (the quicker) programming language C.

**Figure 3.2: Signing Futures Market Trades - Example for 1995/01/03**

This figure illustrates how the State Space Form Regime Switching (*SSFRegSw*) signing algorithm is applied to the data. The observations consist of the sequence of the prices reported on January 3, 1995. If multiple trades are observed in the same second at the same price we consider this to be one observation. In the top plot the first 100 reported prices are indicated with crosses. The smoothed efficient price series obtained using the *SSFRegSw* methodology is given by the solid line. The bottom plot gives the smoothed probability that the trade is initiated by the buying party for the first 100 observations. We label a trade as 'buyer-initiated' if this probability is greater than 0.5.



In general we see our signing methodology acts in the same way as the Lee and Ready (1991) methodology for signing trades when a bid-ask quote is available. In the latter case a trade is considered to be buyer-initiated if it is above the midquote, which is not unlike our case where we label a trade to be initiated by the buying party if it takes place above the smoothed efficient price. Thus in both cases an estimate of the underlying true price is obtained, the midquote for Lee and Ready (1991) and the smoothed efficient price for the SSFRegSw method, around which transactions take place.

Finally, in Panel B of Table 3.2 we provide some statistics of the results of the signing algorithm on a subset of our dataset. The sign we obtain with our SSF Regime Switching methodology agrees with Hasbrouck's method about 87% of all observations, and between 82% and 89% with the SSF Approximating method. Moreover, our concerns with calculation time seem to be justified: for an average day Hasbrouck's methodology requires more than 2 minutes, while our alternative SSF methods require less than 7 seconds.

We obtain a half-spread estimate of about 0.15. Following the calculations in Hasbrouck (2004) we transform this into a dollar figure by multiplying this estimate with the average transaction price. As the transaction price is very roughly about \$110 in 1994-1997, we get a spread estimate of  $0.15 \times 2 \times \$110 = \$33$ . This is very close to the tick size on the market, which is \$31.25 (see [www.cbot.com](http://www.cbot.com)). This adds further support to the reliability of the signing algorithm.

## 3.5 Empirical Results

### 3.5.1 Initiated Trades of the Market Makers

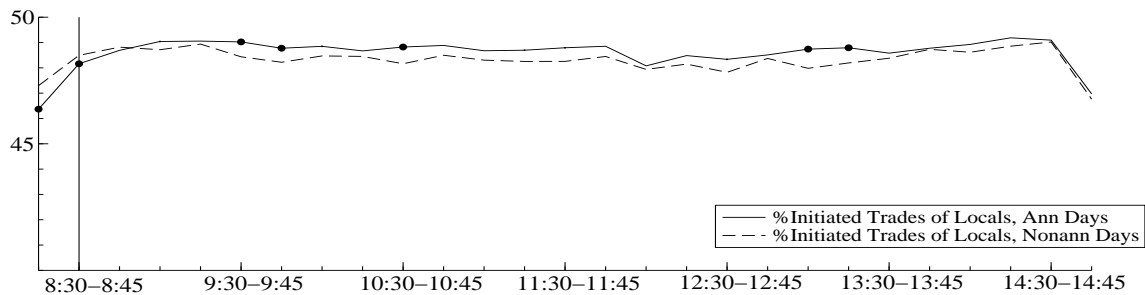
To study the liquidity supply role of market makers we first examine what portion of their trades they initiate. As described in Section 3.3, there are three groups of floor traders active on the 30Y treasury futures market. These are the local traders, who only trade for their own account, the dual traders, who trade for own account and on behalf of customers, and brokers, who only trade on behalf of customers. Of these the first two provide market making services, and these are the groups that we focus on in this study. The difference between these two is, besides possible trader heterogeneous effects, the information set: in addition to the publicly available information the dual traders observe the customer trades they bring to the market, from which they can also make inferences. Due to this we split our results for market makers to locals and duals.

In addition we split our results to those for days with macroeconomic announcements and compare these to days with no macroeconomic announcements. As on the former group of days it is known information will come to the market it is interesting to see if that extra information leads to difference in trade behavior.

**Figure 3.3: Percentage of Trades Initiated by Market Makers**

These figures show the intraday pattern of the percentage of proprietary trades of locals (A) and duals (B) that they initiate. We classify traders at the daily basis, and label a trader to be a local (broker) if more than 98% (less than 2%) of his trades are for own account, otherwise he is a dual. We sign data according to our State Space Form Regime Switching methodology and match the obtained buyer- and seller-initiated indicator with the buy and sell indicator from our dataset. If they agree we classify the trade as being initiated by the market maker. For each day and trader group we calculate the percentage of total proprietary trades that they initiate, we label this as *%initiated trades*. The solid (dashed) lines show the intraday pattern for announcement (nonannouncement) days, the solid vertical lines represent the 8:30-8:45 announcement interval. A circle indicates a significant difference between announcement and nonannouncement days at the 99% level.

(A) %Initiated Trades of Locals



(B) %Initiated Trades of Duals

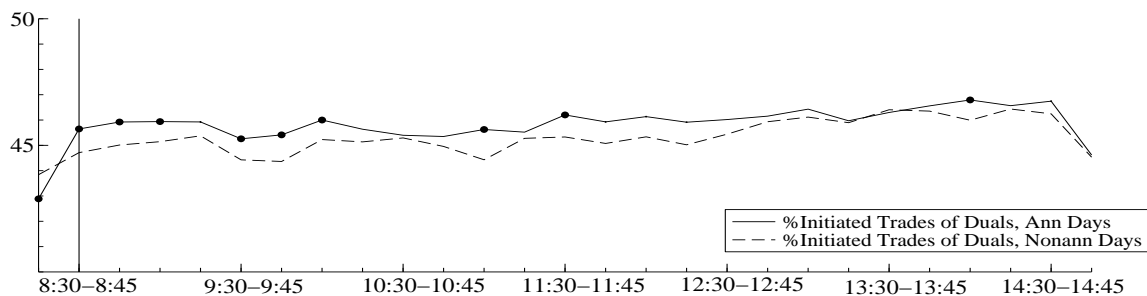


Figure 3.3 shows the intraday pattern of the percentage of the market makers' trades that they initiate. On each day and in each 15 minute interval we calculate the total number of proprietary trades in which market makers are involved. Then we use the buyer- or seller-initiated indicator that we obtain from our State Space Form Regime Switching signing methodology from Section 3.4 and match it to the buy/sell indicator in our dataset. If these agree we label the trade as being initiated by a market maker. We then calculate the number of initiated market maker trades as a percentage of the total number of market maker trades. In Panel (A) we look at this variable for locals, in Panel (B) for duals.

For both locals and duals the percentage of initiated trades is high. On average market makers initiate more than 45% of their trades. As the trades are two-sided the maximum percentage for all market makers we could have here is 50%. That the percentage of initiated trades is close to this number indicates that market makers also demand liquidity for a significant part of the day.

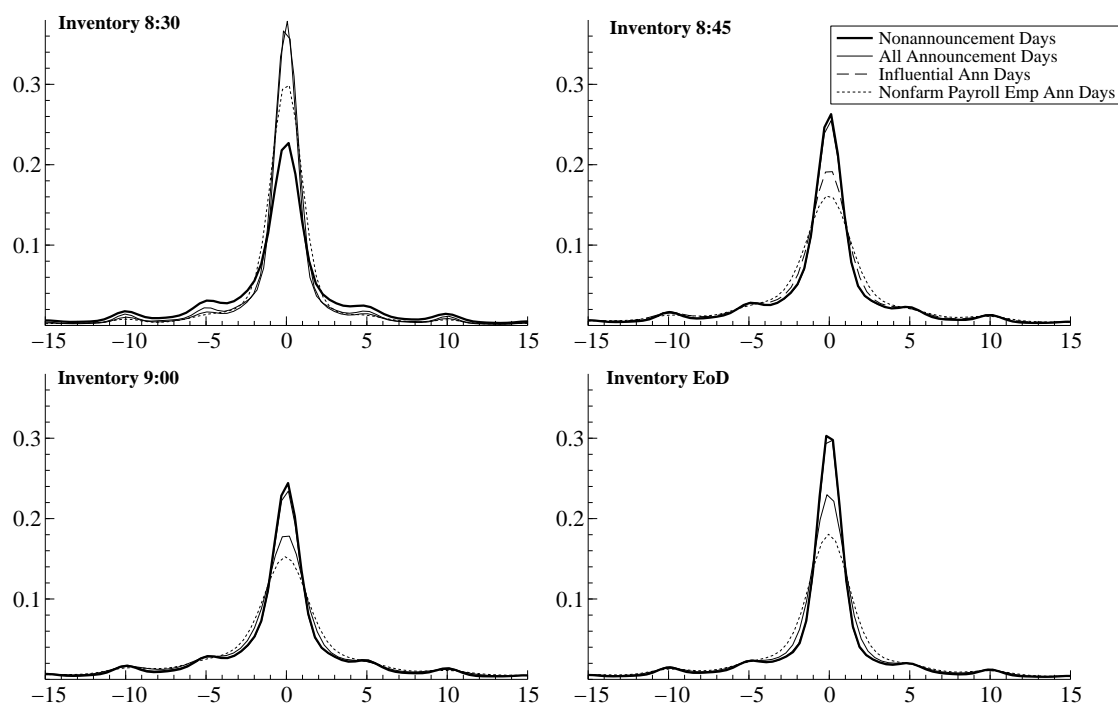
That this percentage is high contradicts the assumption in classic market making models that the market maker is a passive liquidity supplier. However, it is not inconsistent with other economic theories. For example, in the 'hot-potato' trading model of Lyons (1997) in the first stage market makers trade with the general public and absorb their order flow. In the second stage in multiple rounds the market makers offset their inventory position by engaging in interdealer trading. Thus, one outside order brought to the market results in multiple trades on the interdealer market, consistent with a high percentage of initiated market maker trades.

The percentage of initiated trades is higher for local traders than for dual traders, 48% and 45% respectively. Interestingly, the difference in this percentage between days when information is coming to the market compared to nonannouncement days is significant for the dual traders in the post-announcement interval. Locals initiate slightly less trades in the announcement interval itself, but do not show a significant difference in the intervals thereafter.

These results are consistent either with the market makers being active in managing their inventory positions, or with them being active in building up speculative positions. In addition, the significant difference for dual traders in the intervals after macroeconomic announcements indicates they change their behavior shortly after there is information coming to the market. To study these issues further it will be interesting to look at the inventory positions that the market makers build up.

**Figure 3.4: Inventory Positions over Day, Three Types of Announcements**

The figure reports the distribution of inventory positions for different times in the day on three type of announcement days and on nonannouncement days for floor traders active in the 30Y treasury futures in 1994-1997. For each time point the inventory position is calculated only for traders that were active before that time, assuming a zero inventory position at the beginning of the day. On announcement days, the distribution of inventory is calculated separately for three groups of announcements: all announcements (indicated by *All Announcement Days*), Nonfarm payroll employment, CPI and PPI announcements (indicated by *Nonfarm, CPI, PPI Ann Days*) and Nonfarm payroll announcements only (indicated by *Nonfarm Payroll Emp Ann Days*). The distribution is shown for four time points in the day: 8:30, 8:45, 9:00 and the end of the trading day (*EoD*).





### 3.5.2 Market Maker Inventory and Macro News

In Figure 3.4 we show the inventory position, calculated as described in Section 3.3, of the market makers at four points in the day. We focus on the difference in the inventory between announcement and nonannouncement days. As the announcement days we consider are defined to have one or more macroeconomic announcement at 8:30 we show the inventory position around this time. We show the empirical distribution of trader inventories at 8:30, thus shortly *before* the news is released, the inventory position at 8:45, at 9:00 and at the end of the trading day.

From these figures it is clear the market makers prefer to enter the announcement with a zero inventory position, they like to ‘go in flat’. The inventory distribution on nonannouncement days at 8:30 (the bold line) is much more dispersed than the distribution on announcement days. Immediately after the announcement the market makers quickly build up an inventory position: compared to nonannouncement days the empirical inventory distribution is more dispersed.

To illustrate how strong these results are we do not only show these for the set of all announcement days, but also refine them to the set of the three most influential announcement days (the nonfarm, CPI and PPI announcement, see Chapter 2 and, e.g., Green (2004) and Pasquariello and Vega (2007)) and only the most influential announcement (the nonfarm payroll employment figure). We find that on days when there is the strongest reaction to the news the distribution of inventory is widest. In addition we split the results to the inventory distribution of locals and duals (results not shown here). The results for these market maker types show similar patterns, and are very similar to each other.

It remains a question however whether these positions are consistent with market maker liquidity supply behavior or speculative position taking. In the first situation the market makers accommodate the desire of the outside customers to trade after announcements by absorbing their trades and in the process thus accumulate an inventory position. In the second situation the inventory position is build up because the market maker has his own projection on the movement of the price, and wants to speculate on this.

### 3.5.3 Inventory Increasing Trades

To disentangle whether the results so far are consistent with liquidity supply or speculative position taking by market makers it is interesting to see what share of the large percentage of initiated trades that we find in Figure 3.3 increases the inventory position of the market makers. If the large percentage of initiated trades and the increased inventory position after macroeconomic announcements is consistent purely with their market

maker liquidity supply role we expect them to only initiate trades that reduce their absolute inventory position. In other words, the market makers will only initiate trades that bring them closer to their preferred inventory position. On the other hand, if we find a high percentage of initiated trades that increase the market makers' absolute inventory position this is consistent with the market makers being informed and speculating on future price movements.

In Panels (A) and (B) of Figure 3.5 we split the percentage of initiated trades from Figure 3.3 to trades that are inventory increasing and those that are inventory decreasing. On each day and in each 15 minute interval we calculate the total number of proprietary trades in which market makers are involved. Then we use the buyer- or seller-initiated indicator that we obtain from our State Space Form Regime Switching signing methodology from Section 3.4 and match it to the buy/sell indicator in our dataset. If these agree we label the trade as being initiated by a market maker. Then, for each market maker we sum all their signed initiated trades in the interval and see whether it increases their inventory position. We then calculate the number of initiated inventory increasing market maker trades as a percentage of the total number of market maker trades.<sup>23</sup> In Panel (A) we look at this variable for locals, in Panel (B) for duals.

We find a very high percentage of market maker trades that are initiated and increase individual inventory positions, over 25%. Unlike the total percentage of initiated trades, which is higher for local traders, the percentage of initiated inventory increasing trades is almost identical for local and dual traders. Interestingly, for both traders there is a significant different percentage of initiated inventory increasing trades in the 8:30-8:45 interval after announcements compared to nonannouncement days. This effect is strongest for the dual traders.

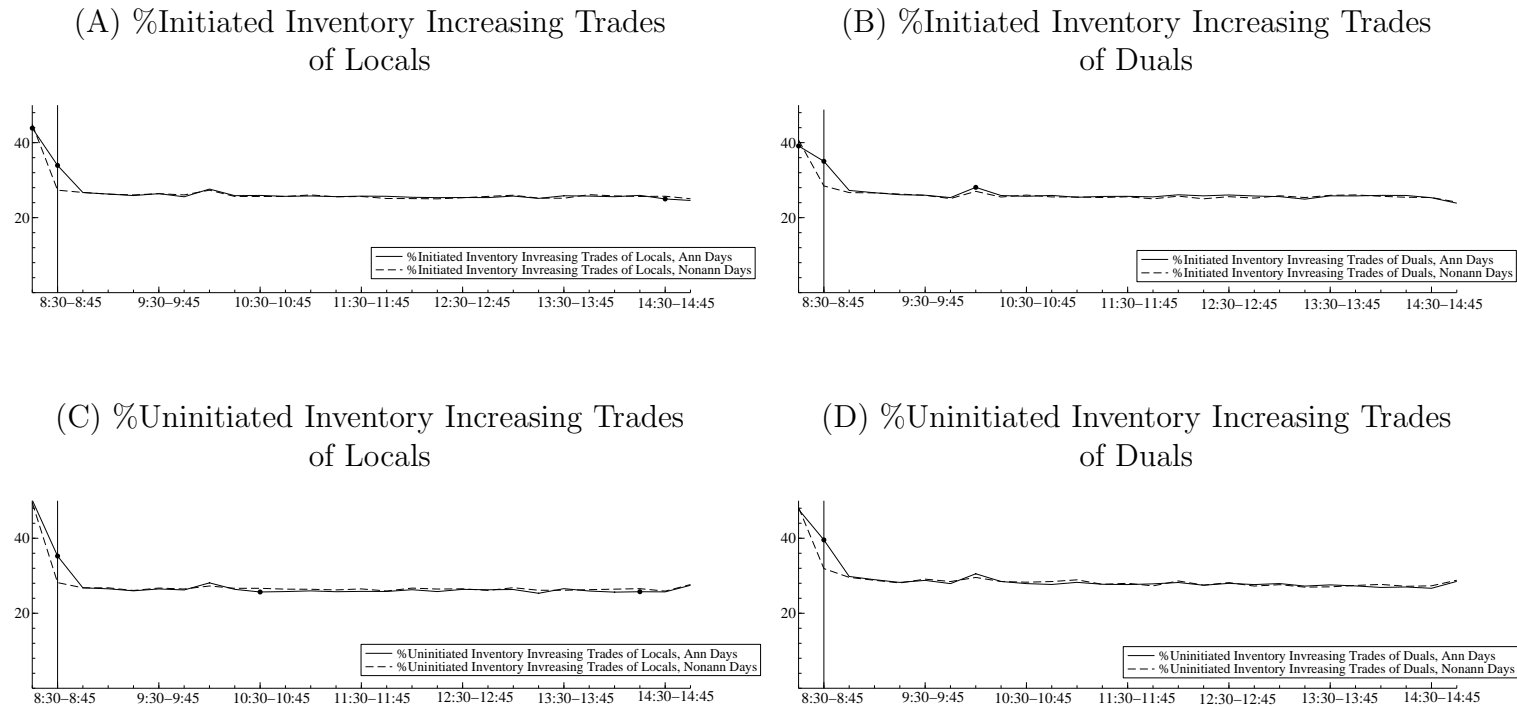
In addition we split the percentage of market maker trades that are uninitiated (the complement of the percentages shown in Figure 3.3) to the part that is inventory increasing (referred to as uninitiated inventory increasing trades) and the part that is inventory decreasing. We show these for the local traders in Panel (C) of Figure 3.5, and in Panel (D) for the dual traders. The results are similar to those for the initiated inventory increasing trades, though here the percentage is slightly higher for dual traders (29% compared to 27%). Most position taking takes place in the 8:30-8:45 announcement interval, and is strongest for the dual traders.

---

<sup>23</sup>The frequency with which we perform this analysis is low. It is very well possible that within the 15 minute interval the initiated trades do not only respond to the inventory position at the begin of the interval, but also to uninitiated trades in the interval. We look at 15 minute intervals to ensure that possible errors due to the timing algorithm do not cause our results. In addition we also perform the analysis of this section in 1 minute intervals. This frequency is consistent with Manaster and Mann (1996), who look at inventory management for a similar futures market dataset. Though the percentages of inventory increasing trades are slightly lower we get qualitatively similar results.

### Figure 3.5: Percentage of Inventory Increasing Trades by Market Makers

These figures show the intraday pattern of the percentage proprietary trades of locals (A, C) and duals (B, D) which increase individual traders' inventory positions. We classify traders at the daily basis, and label a trader to be a local (broker) if more than 98% (less than 2%) of his trades are for own account, otherwise he is a dual. We sign data according to our State Space Form Regime Switching methodology and match the obtained buyer- and seller-initiated indicator with the buy and sell indicator from our dataset. If they agree we classify the trade as being initiated by the market maker, if they do not agree we classify the trade as being an uninitiated market maker trade. For each day and trader group we then calculate the percentage of total proprietary trades that they initiate and which increase individual traders inventory (labelled as *%initiated inventory increasing trades*), and do the same for uninitiated trades (*%uninitiated inventory increasing trades*). Panels (A) and (B) show the percentage of trades initiated by the local and dual trader that increase inventory, panels (C) and (D) the percentage of trades not initiated by the local and dual that increase inventory. The solid (dashed) lines show the intraday pattern for announcement (nonannouncement) days, the solid vertical lines represent the 8:30-8:45 announcement interval. A circle indicates a significant difference between announcement and nonannouncement days at the 99% level.



The results in Figure 3.5 are consistent with the market makers taking on a position after macroeconomic announcements. This is done both actively, through initiated trades, and passively, through uninitiated trades. Both local traders and dual traders do so, but in particular the traders with the additional information from bringing customer orders to the market build up a position after the announcement. The high percentage of initiated inventory increasing trades provides evidence against the market makers being uninformed liquidity suppliers, as pure market makers would only actively engage in trading to offset their inventory position. The high percentage of uninitiated inventory increasing trades indicates that the position taking is not only done through initiating trades, but also through passively participating in trades that steer the inventory position in a certain direction.

### 3.5.4 Inventory Increasing Trades and Trading Profits

If the market makers indeed take on speculative positions through both initiated and uninitiated trades then it is interesting to see if the traders that do so most derive positive profits from this. We examine this by looking in the cross-section of market makers, and relate the percentage of inventory increasing trades to profits from trading. We follow Fishman and Longstaff (1992) and define profitability as:

$$\pi_{kt} = \frac{\left( \sum_{j=1}^{N_{kt}^s} q_{jkt}^s P_{jkt}^s - \sum_{j=1}^{N_{kt}^b} q_{jkt}^b P_{jkt}^b + (\sum_{j=1}^{N_{kt}^b} q_{jkt}^b - \sum_{j=1}^{N_{kt}^s} q_{jkt}^s) REF P_t \right)}{\max(\sum_{j=1}^{N_{kt}^b} q_{jkt}^b, \sum_{j=1}^{N_{kt}^s} q_{jkt}^s)}, \quad (3.4)$$

where  $\pi_{kt}$  is the profit per round-trip contract<sup>24</sup> for intermediary  $k$  on day  $t$ ,  $N_{kt}^b$  ( $N_{kt}^s$ ) is the total number of buys (sells),  $q_{jkt}^b$  ( $q_{jkt}^s$ ) is the quantity of the  $j$ th transaction in terms of number of contracts,  $P_{jkt}^b$  ( $P_{jkt}^s$ ) is the associated price, and  $REF P_t$  is the reference price in day  $t$ . Similar to the calculation of inventory, the profit calculation assumes that the intermediary starts with zero inventory. The end of day position (if any) is liquidated at a reference price  $REF P_t$ , which for our full day results we take to be the daily settlement price.

In Table 3.3 we split these profits from trading according to the percentage of inventory increasing trades. For each day we calculate the 25% quantile, median and the 75% quantile of the percentage of inventory increasing trades and based on this classify the traders and corresponding profits on a daily basis in four groups. In Panel A we split the trading profits based on the percentage of initiated inventory increasing trades. We generally find for both the locals and duals on both announcement and nonannouncement

---

<sup>24</sup>We use a per-contract profit measure to control for trade activity, as locals are more active than duals.

**Table 3.3: Own-Account Trading Profits and Inventory Increasing Trades**

This table reports summary statistics on the cross-sectional distribution of proprietary trading profits split to percentage of trades that increase trader inventory for both locals and duals. We classify traders at the daily basis, and label a trader to be a local (broker) if more than 98% (less than 2%) of his trades are for own account, otherwise he is a dual. To obtain the profits per contract traded round trip for each trader we subtract the value of purchases from the value of sales and add the value of end-of-period inventory (assuming zero inventory at the start). We divide this by the total number of contracts traded to arrive at a profit per contract traded round trip. We split these profits according to percentage of inventory increasing trades. That is, for each day we calculate the 25% quantile ( $Q(25\%)$ ), the median and the 75% quantile ( $Q(75\%)$ ) of the percentage of inventory increasing trades and based on this classify daily the traders and corresponding profits in four groups. In Panel A we split the trading profits based on the %initiated inventory increasing trades (%*iiit*), in Panel B this is done according to the %uninitiated inventory increasing trades (%*uiit*). We show the mean, standard deviation (*St Dev*) and the three quartiles (*25% Quant*, *Median* and *75% Quant*) of the cross-sectional distribution (across intermediaries) of own-account trading profits (with the number of trader days in each group in the column #*Trader Days*).

Panel A: Trading Profits split to %Initiated Inventory Increasing Trades						
	#Trader Days	Mean	St Dev	25% Quant	75% Median	75% Quant
Local Traders						
announcement days						
% <i>iiit</i> < $Q(25\%)$	27,868	5.9	158.1	-9.9	5.1	21.5
$Q(25\%) \leq \%iiit < \text{Median}$	27,748	5.4	67.3	-3.4	5.7	15.4
$\text{Median} \leq \%iiit < Q(75\%)$	27,842	6.7	78.3	-1.3	6.5	15.4
% <i>iiit</i> $\geq Q(75\%)$	28,187	5.1	165.7	-2.7	6.7	19.2
nonannouncement days						
% <i>iiit</i> < $Q(25\%)$	24,028	3.7	133.7	-11.1	4.4	20.2
$Q(25\%) \leq \%iiit < \text{Median}$	24,081	4.9	63.1	-4.2	4.9	14.5
$\text{Median} \leq \%iiit < Q(75\%)$	24,096	6.5	63.4	-2.1	5.7	14.3
% <i>iiit</i> $\geq Q(75\%)$	24,469	6.0	146.6	-4.6	5.5	17.4
Dual Traders						
announcement days						
% <i>iiit</i> < $Q(25\%)$	14,658	5.9	89.7	-9.6	8.1	25.3
$Q(25\%) \leq \%iiit < \text{Median}$	14,624	8.4	59.0	-3.6	9.1	21.4
$\text{Median} \leq \%iiit < Q(75\%)$	14,729	9.3	57.2	-0.7	9.2	19.9
% <i>iiit</i> $\geq Q(75\%)$	15,188	10.3	84.0	-0.0	8.2	19.4
nonannouncement days						
% <i>iiit</i> < $Q(25\%)$	13,025	5.1	79.0	-9.6	7.8	24.0
$Q(25\%) \leq \%iiit < \text{Median}$	12,978	6.7	50.0	-4.1	8.2	19.6
$\text{Median} \leq \%iiit < Q(75\%)$	13,060	8.0	50.2	-1.8	8.3	18.5
% <i>iiit</i> $\geq Q(75\%)$	13,601	5.8	70.5	-1.4	6.8	17.0

(continued on next page)

*(continued from previous page)*

Table 3.3, Panel B: Trading Profits split to %Uninitiated Inventory Increasing Trades						
	#Trader			25%		75%
	Days	Mean	St Dev	Quant	Median	Quant
<b>Local Traders</b>						
announcement days						
%uiit < Q(25%)	27,886	4.7	150.1	-7.7	5.8	22.0
Q(25%) ≤ %uiit < Median	27,728	6.5	65.6	-3.7	5.2	15.0
Median ≤ %uiit < Q(75%)	27,853	4.5	71.7	-2.6	5.5	14.0
%uiit ≥ Q(75%)	28,178	7.3	176.4	-1.8	7.8	20.9
nonannouncement days						
%uiit < Q(25%)	24,081	5.6	133.7	-9.0	5.0	21.5
Q(25%) ≤ %uiit < Median	24,070	5.5	58.3	-4.6	4.5	14.1
Median ≤ %uiit < Q(75%)	24,098	4.9	66.7	-3.4	4.8	13.0
%uiit ≥ Q(75%)	24,425	5.2	147.2	-3.7	6.6	18.2
<b>Dual Traders</b>						
announcement days						
%uiit < Q(25%)	14,701	10.5	87.5	-7.7	8.3	26.8
Q(25%) ≤ %uiit < Median	14,640	7.7	60.4	-4.4	8.0	20.8
Median ≤ %uiit < Q(75%)	14,736	7.4	56.1	-2.0	8.2	19.2
%uiit ≥ Q(75%)	15,122	8.2	86.0	0.0	9.9	20.5
nonannouncement days						
%uiit < Q(25%)	13,057	7.0	73.4	-8.4	7.6	24.3
Q(25%) ≤ %uiit < Median	13,016	6.5	50.3	-5.4	7.0	19.1
Median ≤ %uiit < Q(75%)	13,094	5.9	48.8	-2.9	7.4	17.4
%uiit ≥ Q(75%)	13,497	6.3	77.0	0.0	8.7	18.8

days a positive relation between the percentage of initiated inventory increasing trades and trading profits.<sup>25</sup> Moreover, the results do not only seem to be in the mean and median, also in the lower quantiles similar patterns can be found. Thus not only do market makers that have a large percentage of initiated inventory increasing trades earn higher profits from trading, they reduce the downside of their profits.

In Panel B of Table 3.3 we relate the percentage of uninitiated inventory increasing trades to profits from trading. There is evidence of similar patterns as documented above for initiated inventory increasing trades. However, for both locals and duals the relation seems to be a bit weaker. This could indicate that uninitiated inventory trades are more due to liquidity supply rather than speculation.

To study whether initiated or uninitiated inventory increasing trades has the strongest relation to the trading profits and see whether the above results also hold for the mean profits we perform a regression. We regress the trading profits on both the percentage of initiated inventory increasing trades and the percentage of uninitiated inventory increasing trades and some control variables. The most natural candidates for control variables are volatility and market maker competition, as it is likely these influence trading profits of market makers. In particular, we estimate the following regression:

$$\begin{aligned} \pi_{j,t}^{d,f} = & \alpha^{d,f} + \beta_1^{d,f} IIIIT_{j,t}^{d,f} + \beta_2^{d,f} UIIT_{j,t}^{d,f} \\ & + \beta_3^{d,f} VOLA_t + \beta_4^{d,f} COMP_t + \sum_k I_{k,t} \gamma_k^{d,f} |S_{k,t}| + \varepsilon_{j,t}^{d,f}, \end{aligned} \quad (3.5)$$

where  $\pi_{j,t}^{d,f}$  is trader  $j$ 's own-account profit per round trip on day  $t$ ,  $IIIIT_{j,t}^{d,f}$  and  $UIIT_{j,t}^{d,f}$  measure the percentage inventory increasing trades,  $VOLA_t$  is the volatility,  $COMP_t$  is a competition proxy and is defined as the number of active market makers,  $I_{k,t}$  is a dummy that is one if there is an announcement of type  $k$  on day  $t$  and zero else,  $S_{k,t}$  is the macro surprise, and  $\varepsilon_{j,t}^{d,f}$  is the error term. We estimate the equation for two types of days  $d$ , announcement days ( $d = ad$ ) and nonannouncement days ( $d = nd$ ), and two floor trader types  $f$ , local traders ( $f = lt$ ) and dual traders ( $f = dt$ ). We have two variables that measure the percentage of inventory increasing trades: the percentage initiated inventory increasing trades ( $IIIIT_{j,t}^{d,f}$ ) and the percentage uninitiated inventory increasing trades ( $UIIT_{j,t}^{d,f}$ ). For estimation, we use the Feasible Efficient GMM procedure with a Newey-West estimator (using three lags) for standard errors. The difference in results when comparing the median and mean in Table 3.3 indicate that there possibly are outliers in the profits series. To ensure these outliers do not cause the significance of estimates we set the 5% smallest and largest values at the 5% and 95% quantile respectively.

---

<sup>25</sup>The pattern is not always monotonous, which is in part caused by traders that are very inactive. For example, traders that only trade once on a day are automatically put in either the smallest or largest group, as either 0% or 100% of their trades are initiated inventory increasing.

**Table 3.4: Determinants of Own-Account Trading Profits**

This table reports the estimation results of the following regression:

$$\pi_{j,t}^{d,f} = \alpha^{d,f} + \beta_1^{d,f} IIIIT_{j,t}^{d,f} + \beta_2^{d,f} UIIT_{j,t}^{d,f} + \beta_3^{d,f} VOLA_t + \beta_4^{d,f} COMP_t + \sum_k I_{k,t} \gamma_k^{d,f} |S_{k,t}| + \varepsilon_{j,t}^{d,f}$$

where  $\pi_{j,t}^{d,f}$  is trader  $j$ 's own-account profit per round trip on day  $t$ ,  $IIIIT_{j,t}^{d,f}$  and  $UIIT_{j,t}^{d,f}$  measure the percentage inventory increasing trades,  $VOLA_t$  is the volatility,  $COMP_t$  is a competition proxy and is defined as the number of active market makers,  $I_{k,t}$  is a dummy that is one if there is an announcement of type  $k$  on day  $t$  and zero else,  $S_{k,t}$  is the macro surprise, and  $\varepsilon_{j,t}^{d,f}$  is the error term. We estimate the equation for two types of days  $d$ , announcement days ( $d = ad$ ) and nonannouncement days ( $d = nd$ ), and two floor trader types  $f$ , local traders ( $f = lt$ ) and dual traders ( $f = dt$ ). We have two variables that measure the percentage of inventory increasing trades: the %initiated inventory increasing trades ( $IIIIT_{j,t}^{d,f}$ ) and the %uninitiated inventory increasing trades ( $UIIT_{j,t}^{d,f}$ ). For estimation, we use the Feasible Efficient GMM procedure with a Newey-West estimator (using three lags) for standard errors.

Dependent Variable:			
Trading Profit per Contract Traded Round Trip			
	(1)	(2)	(3)
%Init inv incr trades			
locals			
ann days	-0.0204 -1.51	-0.0188 -1.39	-0.0212 -1.57
nonann days	-0.00962 -0.732	-0.0093 -0.707	-0.00853 -0.649
duals			
ann days	0.0357** 2.83	0.0378** 3	0.0389** 3.09
nonann days	-0.0102 -0.854	-0.00963 -0.807	-0.0104 -0.867
%Uninit inv incr trades			
locals			
ann days	0.0539** 4.3	0.0548** 4.37	0.0536** 4.28
nonann days	0.0267* 2.2	0.027* 2.23	0.0281* 2.32
duals			
ann days	0.00696 0.621	0.00997 0.889	0.014 1.25
nonann days	0.00527 0.486	0.00532 0.49	0.00592 0.543
Controls		yes	
Day dummies			yes

\*/\*\* indicates significance at the 95%/99% level.



Table 3.4 reports the estimates for the above regression. We estimate three variations of (3.5). First, we estimate (3.5) without the control variables volatility, competition and the announcement surprises. We only find significance of the positive relation between initiated inventory increasing trades and trading profits for the dual traders on announcement days. In the second set-up, with control variables, we get the same result. In a third set-up we are careful to take any day specific effects into account by including a dummy for every day.<sup>26</sup> In this set-up the strong results for the dual traders on announcement days are confirmed. Interestingly, across all three models we find a positive and significant relation between uninitiated inventory increasing traders and profits for the local traders.

We conclude that for locals and duals on announcements days there indeed is a positive and significant relation between inventory increasing trades and profits from trading, with the strongest relation for the dual traders. These results are consistent with the market makers building up a position after the announcement, and earning a profit from this. The market makers that have the highest percentage of inventory increasing trades earn the highest profits. Moreover, this relation is strongest for the group of market makers with the additional information set of observing what orders customers bring to the market.

### 3.5.5 Bid-Ask Spread of Market Makers

Our results so far show that the market makers with the largest information set, the dual traders, are most active in the interval after announcements to build up an inventory position compared to nonannouncement days. Moreover, we find that these traders earn the highest profits from trading, and that this is positively related to the percentage of inventory increasing trades.

If these dual traders have more information and the other market participants have a signal on which market makers are dual traders, then these other participants should charge these dual traders a higher spread to protect themselves against this information. In Figure 3.6 we examine this thesis, by comparing the bid-ask spread of dual traders to that of local traders. To calculate the bid-ask spread we first use our State Space Form Regime Switching methodology to obtain a buyer- and seller-initiated indicator. We then follow the Manaster and Mann (1996) methodology to calculate spread in futures markets: we calculate the difference between the average (volume-weighted) buy price and the average sell price.

On both announcement and nonannouncement days duals get charged a higher spread than local traders, \$28 compared to \$26. However, consistent with the duals having a significantly larger percentage of inventory increasing trades and higher profits in the post

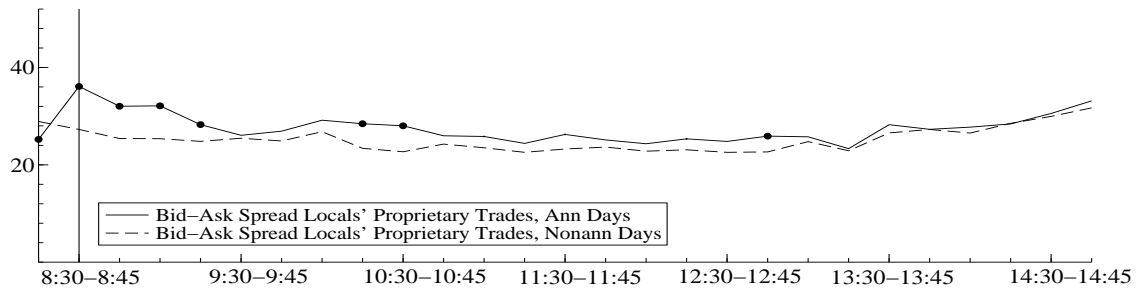
---

<sup>26</sup>Note that in this case the control variables can not be included anymore, as they span the same space as the day dummies.

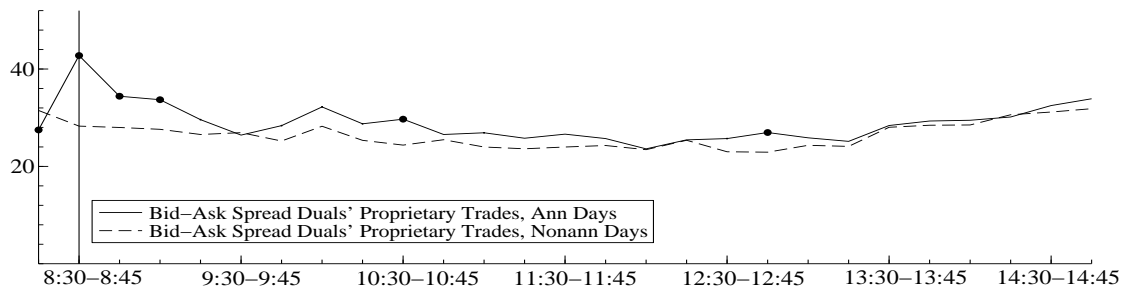
**Figure 3.6: Bid-Ask Spread of Proprietary Order Flow of Locals and Duals**

These figures show the intraday pattern of the Bid-Ask spread calculated from proprietary trades of locals (A) and duals (B) in the 30Y treasury futures market. We classify traders at the daily basis, and label a trader to be a local (broker) if more than 98% (less than 2%) of his trades are for own account, otherwise he is a dual. We sign proprietary trades according to our State Space Form Regime Switching methodology and thus obtain a buyer- and seller-initiated indicator. We then follow the Manaster and Mann (1996) methodology to calculate spread: we calculate the difference between the average (volume-weighted) buy price and the average sell price. The solid (dashed) lines show the intraday pattern for announcement (nonannouncement) days, the solid vertical lines represent the 8:30-8:45 announcement interval. A circle indicates a significant difference between announcement and nonannouncement days at the 99% level.

(A) Bid-Ask Spread (in \$) calculated from Trades of Locals



(B) Bid-Ask Spread (in \$) calculated from Trades of Duals



announcement interval, the bid-ask spread they pay is significantly higher than the spread on nonannouncement days and than the local traders on announcement days.<sup>27</sup>

These results are consistent with the market participants having a signal on who the dual traders are, and charging them a higher spread as they are the most informed market makers. The signal is not perfect however, as we find the dual traders still earn positive profits. Overall our results provide evidence against the market maker being an uninformed passive liquidity supplier.

## 3.6 Conclusion

We examine the 30Y U.S. treasury futures market to study the market maker. Classic models assume he is an uninformed liquidity supplier who actively manages his inventory. So far, the empirical support is poor. We examine a large cross-section of 3,384 intermediaries who to greater or lesser extent provide market making services and study whether position taking by market makers explains the difference between theory and empirics.

We find that market makers initiate a significant amount of trades that increase their inventory positions. When we look at the cross-section of market makers and relate the extent to which their trades are inventory increasing to their profits from trading we find a significant and positive relation. These results are strongest when looking at the high-information environment created by the scheduled releases of macroeconomic news and for market makers that have the largest information set.

Overall our results provide evidence against the market maker being an uninformed liquidity supplier. We document market making behavior consistent with the Madhavan and Smidt (1993) model, in which the market maker is both a dealer and a speculator. Our results stress the need for the development of theoretical models in which the market maker is informed, such as the recent Boulatov and George (2007) model.

In future work we will further investigate the interaction between liquidity supply and liquidity demand. In particular, we will look at whether market makers supply liquidity on some days and demand liquidity on others, or if they do both at the same day. A method to investigate this further is to disentangle the profits from trading. Manaster and Mann (1998) propose a methodology that decomposes total profits from trading into execution

---

<sup>27</sup>The spread we derive here is higher than that calculated from customer trades as in Chapter 2. This finding relates to Dunne, Hau, and Moore (2007), who find higher spread for interdealer trades compared to the customer segment for European bond markets. Also note that the spread we find adds further support to our State Space Form Regime Switching signing algorithm. In the extreme case of the algorithm generating random signs the Manaster and Mann (1996) procedure to calculate spreads would result in a spread of zero dollars. The spread we obtain is significant, and has intraday patterns that agree with those derived from the (already signed) customer trades. Moreover, its average of a little under \$30 agrees with the tick size in the market and the estimates of the half spread which are reported in Section 3.4.2.

and timing profits. Execution profits represent immediate profits from the difference between the trade price and the true price of the asset at that moment. Timing profits can be interpreted as long term speculative profits. When we combine this technique with our signing methodology we are able to identify for each trader how he obtains his profit and further look into the liquidity supply and demand role of market makers.

## Appendix 3A: Recursions of Filter

In this appendix we describe the recursions of the filtering algorithm used to make inferences on the state space form regime switching representation of the Roll (1984) model (see Section 3.4.1). For illustrative purposes we first detail these recursions for a case without regime switching, the Gaussian approximation of the Roll (1984) model (as described in Footnote 19).

### Recursions for Linear State Space Model

The Gaussian approximation of the Roll (1984) model is given by:

$$\begin{aligned} p_t &= m_t + v_t, & v_t &\sim N(0, \sigma_v^2), \\ m_t &= m_{t-1} + u_t, & u_t &\sim N(0, \sigma_u^2), \end{aligned}$$

for  $t = 1, \dots, n$  and with  $p_t$  the observed log price series and  $m_t$  the unobserved log efficient price of the asset. In here  $\sigma_v^2$  and  $\sigma_u^2$  are parameters, with the former representing the spread (in effect approximating the binary  $cq_t$  from the Roll (1984) model with a Gaussian distribution) and the latter the log efficient price innovation; see also Footnote 19. We are interested in obtaining estimates of these parameters and the log efficient price series, conditional on all data. To do this we rely on standard linear Gaussian state space techniques, such as detailed in Durbin and Koopman (2001) and Kim and Nelson (1999). In particular, we use the Kalman filter for this. Following the terminology in these references, the latent  $m_t$  is often referred to as the unobserved ‘state’ of the observed price process.

To initiate the Kalman filter we need to take a distribution for the first efficient price:  $m_1 \sim N(a_1, P_1)$ . In case little is known about this we can make this a very uninformative distribution, by setting  $P_1$  to be large.<sup>28</sup> Then for every time  $t$  we first calculate the estimated unobserved efficient price based on all data up to and including the previous period (which is referred to as the ‘predicted state’). In the second step, we obtain the estimated unobserved efficient price based on all data up to and including this period

---

<sup>28</sup>In the extreme case,  $P_1 \rightarrow \infty$ , the Kalman filter is said to have a diffuse initial distribution.

(referred to as the ‘filtered state’). We have:

$$\begin{aligned} m_{t|t-1} &= \begin{cases} a_1 & t = 1, \\ m_{t-1|t-1} & t > 1, \end{cases} \\ P_{t|t-1} &= \begin{cases} P_1 & t = 1, \\ P_{t-1|t-1} + \sigma_u^2 & t > 1, \end{cases} \\ f_t &= p_t - m_{t|t-1}, \\ F_t &= P_{t|t-1} + \sigma_v^2, \\ m_{t|t} &= m_{t|t-1} + P_{t|t-1} F_t^{-1} f_t \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1} F_t^{-1} P_{t|t-1}, \end{aligned}$$

with  $m_{t|t-1}$  ( $P_{t|t-1}$ ) the predicted state (and its variance),  $m_{t|t}$  ( $P_{t|t}$ ) the filtered state (and its variance) and  $f_t$  ( $F_t$ ) the Kalman filter residual (and its variance). To obtain estimates of the log efficient price conditional on all data (referred to as the ‘smoothed state’) we need to do a backward loop, known as the Kalman Smoother:

$$\begin{aligned} m_{t|n} &= \begin{cases} m_{t|n} & t = n, \\ m_{t|t} + P_{t|t} P_{t+1|t}^{-1} (m_{t+1|n} - m_{t+1|t}) & t < n - 1, \end{cases} \\ P_{t|n} &= \begin{cases} P_{t|n} & t = n, \\ P_{t|t} + P_{t|t} P_{t+1|t}^{-1} (P_{t+1|n} - P_{t+1|t}) P_{t+1|t}^{-1} P_{t|t} & t < n - 1, \end{cases} \end{aligned}$$

with  $m_{t|n}$  ( $P_{t|n}$ ) the smoothed state (and its variance).

As pointed out above, the goal of the analysis is to make inferences about the parameters and the unobserved states. To this end we need to derive the likelihood

$$L(\sigma_v^2, \sigma_u^2 | p_1, \dots, p_n) = f(p_1, \dots, p_n | \sigma_v^2, \sigma_u^2).$$

We use

$$f(p_1, \dots, p_n | \sigma_v^2, \sigma_u^2) = \prod_{t=2}^n f(p_t | p_1, \dots, p_{t-1}, \sigma_v^2, \sigma_u^2) f(p_1 | \sigma_v^2, \sigma_u^2).$$

We follow the standard approach in econometrics to maximize the log of the likelihood rather than the likelihood, such that we obtain:

$$\ln(f(p_1, \dots, p_n | \sigma_v^2, \sigma_u^2)) = \ln(f(p_1 | \sigma_v^2, \sigma_u^2)) + \sum_{t=2}^n \ln(f(p_t | p_1, \dots, p_{t-1}, \sigma_v^2, \sigma_u^2)).$$

We evaluate  $\ln(f(p_t | p_1, \dots, p_{t-1}, \sigma_v^2, \sigma_u^2))$  using the output from the Kalman filter. Fol-

lowing Durbin and Koopman (2001) we obtain:

$$\ln(f(p_t|p_1, \dots, p_{t-1}, \sigma_v^2, \sigma_u^2)) = -\ln(2\pi)/2 - \ln(F_t)/2 + f_t^2 F_t^{-1}/2.$$

From this it is clear that evaluating the loglikelihood through the Kalman filter is straightforward. After initializing the likelihood for the first observation we simply add the above term at every time  $t$ .

The derivation of the Kalman filter and likelihood strongly relies on linearity and normality. In calculating the filtered state  $m_{t|t}$  and its variance  $P_{t|t}$  we use the convenient properties of the bivariate normal distribution. If  $v_t$  and  $u_t$  are not normal the derivation no longer holds and we face a problem.<sup>29</sup> Typically researchers use computationally expensive simulation based techniques, such as Importance Sampling (see Durbin and Koopman (2001, Part II)), to deal with this. However, in the next subsection we discuss an algorithm that can be used when some of the distributions are binary instead of Gaussian. This technique for this special situation has the advantage that it does not rely on simulation based methods.

## Recursions for State Space Model with Regime Switching

The state space form regime switching representation of the Roll (1984) model is:

$$\begin{aligned} p_t &= m_t + cq_t, & q_t &\in \{-1, +1\}, \\ m_t &= m_{t-1} + u_t, & u_t &\sim N(0, \sigma_u^2). \end{aligned}$$

with  $p_t$  the observed log price series,  $m_t$  the unobserved log efficient price of the asset and  $q_t$  the unobserved time series of trade signs. In here  $c$  and  $\sigma_u^2$  are parameters, with the former representing the half-spread and the latter the log efficient price innovation. We are interested in obtaining estimates of these parameters, the log efficient price series and the sign time series, conditional on all data.

As we point out in the previous subsection, to do this we can no longer rely on standard linear Gaussian state space techniques. Fortunately, for the special set-up of the Roll (1984) model we have a likelihood based alternative to the simulation techniques that typically need to be used for non-Gaussian state space models. As discussed in Section 3.4.1, we work from regime switching models in state space form, such as detailed in Kim and Nelson (1999). The associated techniques are frequently used by macroeconomists to model time series which process depends on whether the economy is in the state of

---

<sup>29</sup>In the non-Gaussian case the filtered estimate obtained from the Kalman filter is no longer the minimum mean square estimator. It is however the minimum mean square *linear* estimator, as Duncan and Horn (1972) point out.

recession or expansion. The idea is to integrate over the unobserved binary distributed variables. Conditional on a time series of trade initiating signs  $q_t$ ,  $t = 1, \dots, n$ , the Roll (1984) model is again a linear Gaussian model. There is one complication with this approach, which we elaborate on when we encounter it in the description below. The recursions in this section describing the technique closely follow Kim and Nelson (1999, Chapter 5).

At each point in time we analyze the model in three stages, with an initialization for the first observation. Similar to the linear Gaussian model from the previous section, in the initialization of the system we get  $f(p_1|c, \sigma_u^2)$ . However, as there are two possibilities for the first trade (buyer- or seller-initiated,  $q_1 = +1$  or  $q_1 = -1$  respectively) we need two initial conditions. For both cases we use the same initialization as in the previous section. In addition we need an initial probability that the first trade is buyer-initiated, which we set to be 0.5 following the Roll (1984) model:  $\mathbb{P}(q_1 = +1) = 0.5$ .

In the first stage we evaluate the Kalman filter, with recursions similar to the previous section. For example, for the first three periods we have:

$$\begin{array}{l}
 t = 1 \\
 t = 2 \\
 t = 3
 \end{array}
 \left\{ \begin{array}{l}
 m_{1|0}^i = a_1^i, \\
 f_1^i = p_1 - m_{1|0}^i - q_1 c, \\
 m_{1|1}^i = m_{1|0}^i + P_{1|0}^i (F_1^i)^{-1} f_1^i, \\
 m_{2|1}^{i,j} = m_{1|1}^i, \\
 f_2^{i,j} = p_2 - m_{2|1}^{i,j} - q_2 c, \\
 m_{2|2}^{i,j} = m_{2|1}^{i,j} + P_{2|1}^{i,j} (F_2^{i,j})^{-1} f_2^{i,j}, \\
 m_{3|2}^{i,j,k} = m_{2|2}^{i,j}, \\
 f_3^{i,j,k} = p_3 - m_{3|2}^{i,j,k} - q_3 c, \\
 m_{3|3}^{i,j,k} = m_{3|2}^{i,j,k} + P_{3|2}^{i,j,k} (F_3^{i,j,k})^{-1} f_3^{i,j,k},
 \end{array} \right.
 \begin{array}{l}
 P_{1|0}^i = P_1^i, \\
 F_1^i = P_{1|0}^i, \\
 P_{1|1}^i = P_{1|0}^i - P_{1|0}^i (F_1^i)^{-1} P_{1|0}^i, \\
 P_{2|1}^{i,j} = P_{1|1}^i + \sigma_u^2, \\
 F_2^{i,j} = P_{2|1}^{i,j}, \\
 P_{2|2}^{i,j} = P_{2|1}^{i,j} - P_{2|1}^{i,j} (F_2^{i,j})^{-1} P_{2|1}^{i,j}, \\
 P_{3|2}^{i,j,k} = P_{2|2}^{i,j} + \sigma_u^2, \\
 F_3^{i,j,k} = P_{3|2}^{i,j,k}, \\
 P_{3|3}^{i,j,k} = P_{3|2}^{i,j,k} - P_{3|2}^{i,j,k} (F_3^{i,j,k})^{-1} P_{3|2}^{i,j,k},
 \end{array}
 \right.$$

where  $i \in \{q_1 = +1, q_1 = -1\}$ ,  $j \in \{q_2 = +1, q_2 = -1\}$  and  $k \in \{q_3 = +1, q_3 = -1\}$ . In the Kalman filter we need the previous filtered state to obtain the current predicted state. As we integrate out over the discrete  $q_t$  this will however make the number of possible paths too large. At time  $t$  we do not only need to know  $q_t$  and  $q_{t-1}$ , but also  $q_{t-2}$ ,  $q_{t-3}$ , etc. This means that at time  $t$  there are  $2^t$  possible paths. To overcome this we collapse the filtered state in the state space form regime switching filter. If we do so we do not need to consider the whole past of the trade sign time series  $q_t, q_{t-1}, q_{t-2}, \dots, q_1$ , but only  $q_t$  and  $q_{t-1}$ . With this approximation we follow the recommendation of Kim (1994, p.9), to "... carry at least  $M^{r+1}$  states at each iteration", with  $M$  the number of states and  $r$  the number of lags appearing in the state-space representation.<sup>30</sup>

---

<sup>30</sup>Other approximations are possible. For example, it is possible to collapse the filtered state into one value and thus only record  $q_t$ . Or, instead of only using  $q_t$  and  $q_{t-1}$  we can in addition also consider



Based on this approximation we thus focus on four possible cases in the first stage of the filter: (1) the current and previous trade being buyer-initiated ( $q_t = +1; q_{t-1} = +1$ ); (2) the current trade being buyer-initiated and the previous seller-initiated ( $q_t = +1; q_{t-1} = -1$ ); (3) the current trade being seller-initiated and the previous buyer-initiated ( $q_t = -1; q_{t-1} = +1$ ); and (4) the current and previous trade being seller-initiated ( $q_t = -1; q_{t-1} = -1$ ). Thus, in this first stage we obtain  $m_{t|t-1}^{i,j}$ ,  $P_{t|t-1}^{i,j}$ ,  $f_t^{i,j}$ ,  $F_t^{i,j}$ ,  $K_t^{i,j}$ ,  $m_{t|t}^{i,j}$  and  $P_{t|t}^{i,j}$  for  $i \in \{q_{t-1} = +1, q_{t-1} = -1\}$  and  $j \in \{q_t = +1, q_t = -1\}$ .

In the second stage we look at the probability of each of the states, given our output of the Kalman filter from these four possible cases. This methodology follows the Hamilton (1989) filter. In particular, for the Roll (1984) model we have:

$$\begin{aligned}
\mathbb{P}(q_t, q_{t-1} | p_1, \dots, p_{t-1}) &= \mathbb{P}(q_t | q_{t-1}) \mathbb{P}(q_{t-1} | p_1, \dots, p_{t-1}) = 0.5 \mathbb{P}(q_{t-1} | p_1, \dots, p_{t-1}), \\
f(p_t | p_1, \dots, p_{t-1}) &= \sum_{q_t} \sum_{q_{t-1}} f(p_t, q_t, q_{t-1} | p_1, \dots, p_{t-1}), \\
&= \sum_{q_t} \sum_{q_{t-1}} f(p_t | q_t, q_{t-1}, p_1, \dots, p_{t-1}) \mathbb{P}(q_t, q_{t-1} | p_1, \dots, p_{t-1}), \\
\text{lik} &+ = \ln(f(p_t | p_1, \dots, p_{t-1})), \\
\mathbb{P}(q_t, q_{t-1} | p_1, \dots, p_t) &= \frac{f(p_t, q_t, q_{t-1} | p_1, \dots, p_{t-1})}{f(p_t | p_1, \dots, p_{t-1})}, \\
&= \frac{f(p_t | q_t, q_{t-1}, p_1, \dots, p_{t-1}) \mathbb{P}(q_t, q_{t-1} | p_1, \dots, p_{t-1})}{f(p_t | p_1, \dots, p_{t-1})}, \\
\mathbb{P}(q_t | p_1, \dots, p_t) &= \sum_{q_{t-1}} \mathbb{P}(q_t, q_{t-1} | p_1, \dots, p_t),
\end{aligned}$$

where  $f(p_t | q_t, q_{t-1}, p_1, \dots, p_{t-1})$  can be evaluated from the prediction error decomposition using the output of the Kalman filter at time  $t$ :

$$f(p_t | q_t, q_{t-1}, p_1, \dots, p_{t-1}) = (2\pi)^{-n/2} |F_t^{i,j}|^{-1/2} e^{-(f_t^{i,j})' (F_t^{i,j})^{-1} f_t^{i,j} / 2},$$

which is evaluated for  $i \in \{q_{t-1} = +1, q_{t-1} = -1\}$  and  $j \in \{q_t = +1, q_t = -1\}$ .

Then, in the third stage for each time point, we follow Kim (1994) and do the aforementioned collapsing of the filtered state. As pointed out above, this is done to overcome the problem that at each point in time we have two possible states due to which the total number of paths will rapidly increase: at time  $t$  there are  $2^t$  possibilities. This is the one approximation we need to keep the filtering analysis tractable. We collapse the filter as

---

$q_{t-2}$ . However, according to Kim's recommendation we are already rather conservative: if we follow his recommendation we should consider only 2 states whereas we look at 4 states.

follows:

$$m_{t|t}^j = \frac{\sum_i \mathbb{P}(q_t = j, q_{t-1} = i | p_1, \dots, p_t) m_{t|t}^{i,j}}{\mathbb{P}(q_t = j | p_1, \dots, p_t)},$$

$$P_{t|t}^j = \frac{\sum_i \mathbb{P}(q_t = j, q_{t-1} = i | p_1, \dots, p_t) (P_{t|t}^{i,j} + (m_{t|t}^j - m_{t|t}^{i,j})^2)}{\mathbb{P}(q_t = j | p_1, \dots, p_t)}.$$

Note that we indeed reduce the dimensionality, as from the Kalman filter we obtain  $m_{t|t}^{i,j}$  and  $P_{t|t}^{i,j}$  for  $i \in \{q_{t-1} = +1, q_{t-1} = -1\}$  and  $j \in \{q_t = +1, q_t = -1\}$ , whereas we now have  $m_{t|t}^j$  and  $P_{t|t}^j$  for  $j \in \{q_t = +1, q_t = -1\}$ . In effect the variable  $m_{t|t}^j$  is a mixture of normals. As a result of the approximation  $m_{t|t}^{i,j}$  can be interpreted as the linear projection of  $m_t$  on  $p_t$  and  $m_{t-1|t-1}^i$ .

The procedure above details the calculations in each of the three stages for each time point in the forward loop of the filter. To obtain estimates of the unobserved log efficient price and trade sign time series conditional on all observations we need to follow these with the backward recursions for the smoother. This also consists of three stages, which differ slightly from those of the filtering algorithm. First, we calculate the smoothed probability for being in a certain state:

$$\mathbb{P}(q_t = j, q_{t+1} = k | p_1, \dots, p_n) = \frac{\mathbb{P}(q_{t+1} = k | p_1, \dots, p_n) \cdot \mathbb{P}(q_t = j | p_1, \dots, p_t) \cdot 0.5}{\mathbb{P}(q_{t+1} = k | p_1, \dots, p_t)},$$

$$\mathbb{P}(q_t = j | p_1, \dots, p_n) = \sum_k \mathbb{P}(q_t = j, q_{t+1} = k | p_1, \dots, p_n).$$

Next, we use the smoothing recursion similar to that of the linear Gaussian model to obtain smoothed values for the efficient price (and the variance thereof). There are again four cases, now depending on the current state (at time  $t$ ) and the next state (time  $t+1$ ). When we obtain  $m_{t|n}^{j,k}$  and  $P_{t|n}^{j,k}$  for  $j \in \{q_t = +1, q_t = -1\}$  and  $k \in \{q_{t+1} = +1, q_{t+1} = -1\}$  we use the state probabilities to obtain smoothed values for the unobserved log efficient price conditional on it being in a certain state:

$$m_{t|n}^j = \frac{\sum_k \mathbb{P}(q_t = j, q_{t+1} = k | p_1, \dots, p_n) m_{t|n}^{j,k}}{\mathbb{P}(q_t = j | p_1, \dots, p_n)},$$

$$P_{t|n}^j = \frac{\sum_k \mathbb{P}(q_t = j, q_{t+1} = k | p_1, \dots, p_n) (P_{t|n}^{j,k} + (m_{t|n}^j - m_{t|n}^{j,k})^2)}{\mathbb{P}(q_t = j | p_1, \dots, p_n)}.$$

Finally, we obtain an estimate of the unobserved log efficient price that is not conditional

on being in a certain state:

$$m_{t|n} = \sum_j \mathbb{P}(q_t = j | p_1, \dots, p_n) m_{t|n}^j.$$

## Chapter 4

# Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters

This chapter is based on Koopman, Mallee, and Van der Wel (2007), which is forthcoming in *Journal of Business and Economic Statistics*.

### 4.1 Introduction

Fitting and predicting time-series of a cross-section of yields has proven to be a challenging task. As with many topics in empirical economic analysis there is the trade-off between the goodness of fit that is obtained by employing statistical models without a reference to economic theory, and the lack of fit by economic models that do provide a basis for the underlying economic theory.

For many decades work on the term structure of interest rates has mainly been theoretical in nature. In the early years work focused on the class of affine term structure models, see Vasicek (1977) and Cox, Ingersoll, and Ross (1985). Duffie and Kan (1996) generalized the literature and Dai and Singleton (2000) characterized the set of admissible and identifiable models. Another class of models focused on fitting the term structure at a given point in time to ensure no arbitrage opportunities exist, see Hull and White (1990) and Heath, Jarrow, and Morton (1992). It has been shown that the forecasts obtained using the first class of models do not outperform the random walk forecasts, see for example Duffee (2002). The second class of models are not appropriate for forecasting given its focus on the cross-section dimension of yields without a reference to the time

series dimension. Time series models aim to describe the dynamical properties and are therefore more suited for forecasting. This may partly explain the renewed interest in statistical time series models for yield curves.

The papers of Diebold and Li (2006, DL) and Diebold, Rudebusch, and Aruoba (2006, DRA) have shifted attention back to the Nelson and Siegel (1987) model. DL and DRA introduce the dynamic Nelson-Siegel model as a statistical three factor model to describe the yield curve over time. The three factors represent level, slope and curvature of the yield curve and thus carry some level of economical interpretation. More importantly, DL and DRA show that the model-based forecasts outperform many other models including standard time series models such as vector autoregressive models and dynamic error-correction models. In DRA, the Nelson-Siegel framework is extended to include non-latent factors such as inflation. Further they frame the Nelson-Siegel model into a state space model where the three factors are treated as unobserved processes and modeled by vector autoregressive processes. A wide range of statistical methods associated with the state space model can be exploited for maximum likelihood estimation and signal extraction, see Durbin and Koopman (2001). We will follow this approach in which the state space representation of the Nelson and Siegel (1987) model plays a central role.

Parameter estimation in DL and DRA relies on two simplifying assumptions. First, the factor loadings in the Nelson-Siegel model depend on a single loading parameter. To enable the estimation of time-varying latent factors in a linear setting, the factor loadings are kept constant over time for each maturity. In the original Nelson and Siegel (1987) article, the loading parameter is time-varying. In DL, the loading parameter is restricted as constant to keep the factor loadings constant. Second, volatility is kept constant for each maturity and over the full sample period.

We contribute to the literature by introducing time-varying factor loadings and time-varying volatility in the dynamic Nelson-Siegel model. The loading parameter determines all factor loadings in the Nelson-Siegel model. When we allow the loading parameter to be time-varying, all loadings will be time-varying. First, the factor loadings and a common volatility component are allowed to change gradually over time via flexible cubic spline functions. It provides initial evidence that model fit improves when considering time-varying parameters. A disadvantage of spline methods is that different number of knots and knot positions will oftentimes lead to different spline functions and possibly different conclusions with respect to the time-varying nature of the yield curve. Therefore we consider the loading parameter as a stochastically time-varying latent process and we treat it as the fourth factor. The latent factors level, slope and curvature together with the time-varying loading parameter are in this case modeled jointly as a vector autoregressive process. The loading parameter is not a linear function of the observation vector and

therefore we obtain a nonlinear dynamic model. We will show that the nonlinear features in the dynamic Nelson-Siegel model can be treated using extended Kalman filter methods. Next, we introduce time-varying volatility by specifying the common variance as the well-known generalized autoregressive conditional heteroskedasticity (GARCH) process, see Bollerslev (1986). In empirical work it is found that during high volatility periods, the yields for all maturities are highly volatile although some maturities are more volatile than others. The time-varying volatility is introduced for different components of the model. We finally calculate likelihood-ratio values and we present goodness-of-fit statistics to assess the most appropriate time-varying specifications in the dynamic Nelson-Siegel model.

The introduction of time-varying parameters may also shed some light on more recent developments in the term structure literature. The dynamic Nelson-Siegel model does not rely on theoretical concepts such as the absence of arbitrage, see also the discussion in Ang and Piazzesi (2003). Recently, Christensen, Diebold, and Rudebusch (2007) have modified the Nelson-Siegel framework to impose the arbitrage-free condition. As a result, a new class of affine dynamic term structure models is defined. An important condition for the riskfree rate to exist in this framework is that loadings are constant over time. This condition may be validated by allowing the factor loadings to be time-varying as we do in this chapter. Also in the work of Diebold, Li, and Yue (2007) on the global yield curve, constant factors are an important condition.

A third contribution to the literature is that we show how easily the Nelson and Siegel (1987) model in state space form treats missing observations. This is a general property of state space models, but has not yet been explored in this context. Besides the standard unsmoothed Fama-Bliss monthly yields dataset for the period 1972-2000 (as used by DRA and others), we also estimate the dynamic Nelson-Siegel model for U.S. Treasury yields over the period January 1972 up to June 2007 obtained from the Federal Reserve Economic Database (FRED). The latter dataset is interesting as it has more recent data, but can not easily be used in the OLS framework due to its many missing values. We show that in the state space framework unbalanced datasets can be treated in a straightforward manner. In particular, by combining the two datasets we show how well the smoothed values for the missing data approximates the true value. Using the state space framework allows us to include the longest maturity bond (maturing in 30 years), which was not issued during the period February 2002 until January 2006.

There are a number of papers that extend the work of DL and DRA for the Nelson-Siegel model. Bianchi, Mumtaz, and Surico (2006) allow for time-varying variances for the latent factors level, slope and curvature. In this specification, it is implied that the factor loadings for the term-structure are also appropriate weights for the volatility in the

term-structure. This appears to be a strong assumption that needs to be validated. We therefore introduce for each yield in the observation equation, a different factor loading for the common disturbance factor that is subject to a GARCH process. We then examine whether the common volatility factor loadings are a linear combination of the loadings for the factors level, slope and curvature. Yu and Zivot (2007) extend the Nelson-Siegel framework by including corporate bonds. De Pooter (2007) examines the dynamic NS model that is extended by additional factors. It is shown that such extensions can improve both the in-sample fit and the post-sample forecasting performance. De Pooter, Ravazzolo, and Van Dijk (2007) study the forecasting ability of the Nelson-Siegel model by focusing on the predictive gains that can be obtained when macroeconomic variables are included and forecasts of different model specifications are combined. Without adopting the Nelson-Siegel framework, Bowsher and Meeks (2008) introduce a 5-factor model where spline functions are used to model the yield curve and where the knots for these splines act as factors. While their approach allows for a more flexible yield curve some economic intuition of the factors is lost. Moreover, also in this framework, volatility is kept fixed over time.

The remainder of the chapter is organized as follows. Section 4.2 describes the baseline dynamic Nelson-Siegel model and Section 4.3 discusses our new extensions. In Section 4.4 we present, discuss and compare estimation results for different model specifications. Section 4.5 presents an illustration for an unbalanced dataset. Section 4.6 concludes.

## 4.2 The Dynamic Nelson-Siegel Model

In this section we introduce the latent factor model that Nelson and Siegel (1987) have developed for the yield curve. We focus on the model that is slightly adjusted in terms of factorization by Diebold and Li (2006) and is extended here to allow for time-varying parameters. We further discuss the state space approach for this initial extension of the model.

### 4.2.1 The Nelson-Siegel Model

Interest rates are denoted by  $y_t(\tau)$  at time  $t$  and maturity  $\tau$ . For a given time  $t$ , the yield curve  $\theta_t(\tau)$  is some smooth function representing the interest rates (yields) as a function of maturity  $\tau$ . A parsimonious functional description of the yield curve is proposed by Nelson and Siegel (1987). The Nelson-Siegel formulation of the yield is modified by Diebold and Li (2006) to lower the coherence between the components of the yield curve.

The Diebold and Li (2006) formulation is given by

$$\theta_t(\tau) = \theta(\tau; \lambda, \beta_t) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \quad (4.1)$$

where  $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$ , for given time  $t$ , maturity  $\tau$  and fixed coefficient  $\lambda$  that determines the exponential decay of the second and third component in (4.1).

The shape and form of the yield curve is determined by the three components and their associated weights in  $\beta_t$ . The first component takes the value 1 (constant) and can therefore be interpreted as the overall level that influences equally the short and long term interest rates. The second component converges to one as  $\tau \downarrow 0$  and converges to zero as  $\tau \rightarrow \infty$  for a given  $\lambda$ . Hence this component mostly influences short-term interest rates. The third component converges to zero as  $\tau \downarrow 0$  and as  $\tau \rightarrow \infty$  but is concave in  $\tau$ , for a given  $\lambda$ . This component is therefore associated with medium-term interest rates.

Since the first component is the only one that equals one as  $\tau \rightarrow \infty$ , its corresponding  $\beta_{1t}$  coefficient is usually linked with the long-term interest rate. By defining the slope of the yield curve as  $\theta_t(\infty) - \theta_t(0)$ , it is easy to verify that the slope converges to  $-\beta_{2t}$  for a given  $t$ . Finally, the shape of the yield can be defined by  $[\theta_t(\tau^*) - \theta_t(0)] - [\theta_t(\infty) - \theta_t(\tau^*)]$  for a medium maturation  $\tau^*$ , say, two years, and for a given  $t$ . It can be shown that this shape approximately equals  $\beta_{3t}$ .

In case we observe a series of interest rates  $y_t(\tau_i)$  for a set of  $N$  different maturities  $\tau_1 < \dots < \tau_N$  available at a given time  $t$ , we can estimate the yield curve by the simple regression model

$$\begin{aligned} y_t(\tau_i) &= \theta_t(\tau_i) + \varepsilon_{it} \\ &= \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right) + \varepsilon_{it}, \end{aligned} \quad (4.2)$$

for  $i = 1, \dots, N$ . The disturbances  $\varepsilon_{1t}, \dots, \varepsilon_{Nt}$  are assumed to be independent with mean zero and constant variance  $\sigma_t^2$  for a given time  $t$ . The least squares method provides estimates for the  $\beta_{jt}$  coefficients  $j = 1, 2, 3$ . These cross-section estimates can be obtained as long as sufficient interest rates for different maturities are available at time  $t$ .

The series of regression estimates for  $\beta_t$ , for all time periods  $t = 1, \dots, T$ , appear to be strongly correlated over time. In other words, the coefficients are forecastable and hence the Nelson-Siegel framework can be used for forecasting in this way. This has been recognized by Diebold and Li (2006) who implemented the following two-step procedure: first, estimate the  $\beta_t$  by cross-section least squares for each  $t$ ; second, treat these estimates as three time series and apply time series methods for forecasting  $\beta_t$  and hence the yield curve  $\theta(\tau; \lambda, \beta_t)$ .



Diebold and Li (2006) compare their two-step forecasts with those from univariate and multivariate time series methods. The different methods produce similar results. Nevertheless, the two-step forecasting approach does better than forecasting the different interest rates series directly, especially for the longer maturities.

## 4.2.2 The Dynamics of the Latent Factors

Diebold, Rudebusch, and Aruoba (2006) go a step further by recognizing that the Nelson-Siegel framework can be represented as a state space model when treating  $\beta_t$  as a latent vector. For this purpose, the regression equation (4.2) is rewritten by

$$y_t = \Lambda(\lambda)\beta_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \Sigma_\varepsilon), \quad t = 1, \dots, T, \quad (4.3)$$

with observation vector  $y_t = [y_t(\tau_1), \dots, y_t(\tau_N)]'$ , disturbance vector  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$  and  $N \times 3$  factor loading matrix  $\Lambda(\lambda)$  where its  $(i, j)$  element is given by

$$\Lambda_{ij}(\lambda) = \begin{cases} 1, & j = 1, \\ (1 - e^{-\lambda \cdot \tau_i}) / \lambda \cdot \tau_i, & j = 2, \\ (1 - e^{-\lambda \cdot \tau_i} - \lambda \cdot \tau_i e^{-\lambda \cdot \tau_i}) / \lambda \cdot \tau_i, & j = 3. \end{cases}$$

The time series process for the  $3 \times 1$  vector  $\beta_t$  can be modeled by the vector autoregressive (VAR) process

$$\beta_{t+1} = (I - \Phi)\mu + \Phi\beta_t + \eta_t, \quad \eta_t \sim NID(0, \Sigma_\eta), \quad (4.4)$$

for  $t = 1, \dots, n$ , with mean vector  $\mu$  and initial condition  $\beta_1 \sim N(\mu, \Sigma_\beta)$  where coefficient matrix  $\Phi$  and variance matrix  $\Sigma_\beta$  are chosen such that  $\Sigma_\beta - \Phi\Sigma_\beta\Phi' = \Sigma_\eta$  and stationarity of the VAR process must be ensured, see Ansley and Kohn (1986) for an effective reparameterisation. We refer to model (4.3) and (4.4) as the dynamic Nelson-Siegel (DNS) model.

## 4.2.3 Estimation Based on the Kalman Filter

We consider the DNS model (4.3) and (4.4) as a linear Gaussian state space model. The vector of unobserved factors  $\beta_t$  is the state vector and can be estimated conditional on the past and concurrent observations  $y_1, \dots, y_t$  via the Kalman filter. Define  $b_{t|s}$  as the minimum mean square linear estimator (MMSLE) of  $\beta_t$  given  $y_1, \dots, y_s$  with mean square error (MSE) matrix  $B_{t|s}$ , for  $s = t-1, t$ . For given values of  $b_{t|t-1}$  and  $B_{t|t-1}$ , the Kalman filter first computes  $b_{t|t}$  and  $B_{t|t}$ , when observation  $y_t$  is available, using the filtering step

$$b_{t|t} = b_{t|t-1} + B_{t|t-1}\Lambda(\lambda)'F_t^{-1}v_t, \quad B_{t|t} = B_{t|t-1} - B_{t|t-1}\Lambda(\lambda)'F_t^{-1}\Lambda(\lambda)B_{t|t-1}, \quad (4.5)$$

where  $v_t = y_t - \Lambda(\lambda)b_{t|t-1}$  is the prediction error vector and  $F_t = \Lambda(\lambda)B_{t|t-1}\Lambda(\lambda)' + \Sigma_\varepsilon$  is the prediction error variance matrix. The MMSLE of the state vector for the next period  $t + 1$ , conditional on  $y_1, \dots, y_t$ , is given by the prediction step

$$b_{t+1|t} = (I - \Phi)\mu + \Phi b_{t|t}, \quad B_{t+1|t} = \Phi B_{t|t} \Phi' + \Sigma_\eta. \quad (4.6)$$

For a given time series  $y_1, \dots, y_T$ , the Kalman filter computations are carried out recursively for  $t = 1, \dots, T$  with initializations  $b_{1|0} = \mu$  and  $B_{1|0} = \Sigma_\beta$  where  $\Sigma_\beta$  is defined below (4.4). The parameters in the VAR coefficient matrix  $\Phi$ , variance matrices  $\Sigma_\eta$  and  $\Sigma_\varepsilon$  together with  $\mu$  and  $\lambda$  are treated as unknown coefficients which are collected in the parameter vector  $\psi$ . Estimation of  $\psi$  is based on the numerical maximization of the log-likelihood function that is constructed via the prediction error decomposition and given by

$$\ell(\psi) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T v_t' F_t^{-1} v_t. \quad (4.7)$$

As a result,  $\ell(\psi)$  can be evaluated by the Kalman filter for a given value of  $\psi$ . A quasi-Newton optimization method is employed for the purpose of maximization based on the numerical evaluation of the score function. We have implemented the necessary calculations in the matrix programming language `0x` of Doornik (1998) with the use of the `SsfPack` state space functions developed by Koopman, Shephard, and Doornik (1999). A textbook treatment of Kalman filter methods is given by Durbin and Koopman (2001).

The state space framework allows for different dynamic processes for vector  $\beta_t$  in the dynamic Nelson-Siegel model. Also, variance matrices  $\Sigma_\varepsilon$  and  $\Sigma_\eta$  can be full or diagonal. Diebold, Rudebusch, and Aruoba (2006) assume that  $\Sigma_\varepsilon$  is diagonal so that the equations for the different yield maturities are uncorrelated, given  $\beta_t$ . This assumption is often used to reduce the number of coefficients and to obtain computational tractability.

### 4.3 DNS Model with Time-Varying Parameters

In this section we extend the DNS model by making the loading parameter and volatility time-varying. For both extensions we first introduce a spline to see whether there indeed are time-varying components in the model. Next, we treat the loading parameter as a stochastically time-varying latent factor and we introduce time-varying volatility in the variance specification of the disturbances. While the splines can be estimated in the linear framework of the previous section these last two extensions introduce nonlinearities in the model that we will handle by the extended Kalman filter discussed below.

### 4.3.1 Time-Varying Loading Parameter

In the DNS model, the loading parameter  $\lambda$  determines the shape of the yield curve. In the earlier studies, the default is to pre-fix a value for  $\lambda$  without estimation. For example, Diebold and Li (2006) fix  $\lambda$  at 0.0609 while Diebold, Rudebusch, and Aruoba (2006) estimate  $\lambda$  to be 0.077. Yu and Zivot (2007) adopt these values for  $\lambda$  in their empirical study concerning corporate bonds. These studies argue that the loadings  $\Lambda_{ij}(\lambda)$  are not very sensitive to different values of  $\lambda$  as can be illustrated graphically. Hence they argue that  $\lambda$  can be fixed such that it maximizes the loading on the curvature component at some medium term (that is, 30 months for  $\lambda = 0.0609$  and 23.3 months for  $\lambda = 0.077$ ).

Here we emphasize that the estimation of  $\lambda$  is straightforward in a state space framework as it can be included in the parameter vector  $\psi$ , see Section 4.2.3. Keeping  $\lambda$  fixed over the full sample period may be too restrictive as the data usually spans over a long time period. In particular, the maturity at which the curvature factor  $\beta_{3t}$  is maximized and the speed of decay of the slope factor  $\beta_{2t}$  depend only on  $\lambda$  and are fixed as a result. However, these characteristics of the yield curve may have changed over time. The importance of  $\lambda$  and its constancy over time is also discussed in Christensen, Diebold, and Rudebusch (2007) where an arbitrage-free version of the Nelson-Siegel framework is proposed. Given the importance of  $\lambda$ , we study its role in more detail by considering possible changes of  $\lambda$  over time.

We consider two different specifications for a time-varying  $\lambda$ : (i) a cubic spline function and (ii) adding it as a fourth latent factor. In case of a spline function, we let  $\lambda$  depend on some function of time  $t$ , that is

$$\lambda_t = f(t; \lambda^*), \quad t = 1, \dots, n,$$

where  $f(t; \lambda^*)$  is a smooth function depending on time  $t$  and  $k_\lambda \times 1$  vector of coefficients  $\lambda^*$ . We consider a cubic spline function for  $f(t; \lambda^*)$  as formulated by Poirier (1976) and given by

$$\lambda_t = f(t; \lambda^*) = w_t' \lambda^*, \quad t = 1, \dots, n, \quad (4.8)$$

where  $w_t$  is the  $k \times 1$  vector of interpolating weights which are implicitly determined through smoothness conditions for a set of pre-defined knot points. The knot points can be regarded as break-points in the cubic spline function.

The second specification we propose is to treat the time-varying loading parameter  $\lambda_t$  as the latent fourth factor  $\beta_{4t}$  and to include it into the vector  $\beta_t$  below (4.1). The new  $4 \times 1$  vector  $\beta_t$  is then modeled by the VAR process (4.4). The loading matrix  $\Lambda(\lambda)$  in the observation equation (4.3) is then replaced by  $\Lambda(\lambda_t)$  with  $\lambda_t = \beta_{4t}$ . The new observation equation is nonlinear in  $\beta_t$  as we obtain  $y_t = \Lambda(\beta_{4t}) \cdot (\beta_{1t}, \beta_{2t}, \beta_{3t})' + \varepsilon_t$ . This specification

is of particular interest since it allows dynamic interactions between changes in cross-sectional (or cross-maturity) dependence, through  $\beta_{4t}$ , and time series dependence of the yields, through  $\beta_{1t}$ ,  $\beta_{2t}$  and  $\beta_{3t}$ . The resulting model with time-varying loadings will be referred to as the DNS–TVL model. Other dynamic processes for  $\lambda_t$  can be considered including a random walk process.

### 4.3.2 Time-Varying Volatility

Another key aspect in the analysis of the term structure is the recognition that interest rates are the result of trading at financial markets. The volatility in the series may therefore have changed over time as well. In most empirical work on the yield curve, monthly time series of interest rates are analyzed under the assumption that the volatility in the time series is constant over time. A few exceptions are Engle, Ng, and Rothschild (1990) and Bianchi, Mumtaz, and Surico (2006). However, investigating time-varying volatility in the context of the DNS model is a novelty. Although the changes in the volatilities for the different maturities have different intensities, they appear to occur at the same time. For this reason, we focus mainly on a common pattern of time-varying volatility in interest rates.

Here we modify the DNS model by introducing time-varying variance matrices via a common volatility component that is modeled by (i) a spline and (ii) a GARCH process. The first specification is based on a cubic spline function for the time-varying log-variance component common to all maturities, that is  $h_t = h(t; \kappa^*)$  where  $h(t; \kappa^*)$  is defined similarly as  $f(t; \kappa^*)$  in (4.8) and is based on pre-defined knot points and a  $k_\kappa \times 1$  vector of coefficients  $\kappa^*$ . In case of the diagonal variance matrix  $\Sigma_\varepsilon = \Sigma_\varepsilon(h_t)$ , its  $i$ th diagonal element is given by

$$\Sigma_{\varepsilon,ii}(h_t) = \sigma_{\varepsilon,i}^2 + \kappa_i^+ \exp(h_t), \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (4.9)$$

The additional coefficients for this spline specification are collected in the parameter vector  $\kappa = (\kappa_1^+, \dots, \kappa_N^+, \kappa^{*'})'$  where all values are non-negative.

Second, we adopt the common GARCH specification, or the one-factor GARCH model, of Harvey, Ruiz, and Sentana (1992) to introduce a time-varying variance for the disturbances in the observation equation. In particular, we consider the decomposition of the disturbance vector  $\varepsilon_t$  given by

$$\varepsilon_t = \Gamma \varepsilon_t^* + \varepsilon_t^+, \quad t = 1, \dots, T,$$

where  $\Gamma$  is redefined here as a  $N \times 1$  loading vector,  $\varepsilon_t^*$  is a scalar disturbance and  $\varepsilon_t^+$  is a  $N \times 1$  disturbance vector. For identification purposes, vector  $\Gamma$  can be normalized such

that  $\Gamma\Gamma = 1$ . The disturbance components are mutually independent of each other and their distributions are given by

$$\varepsilon_t^* \sim NID(0, h_t), \quad \varepsilon_t^+ \sim NID(0, \Sigma_\varepsilon^+), \quad t = 1, \dots, T, \quad (4.10)$$

where  $\Sigma_\varepsilon^+$  is typically, but not necessarily, a diagonal matrix and where variance  $h_t$  is specified as the GARCH process developed by Bollerslev (1986). In particular, we have

$$h_{t+1} = \gamma_0 + \gamma_1 \varepsilon_t^{*2} + \gamma_2 h_t, \quad t = 1, \dots, T, \quad (4.11)$$

with unknown coefficients  $\gamma_0 > 0$ ,  $0 < \gamma_1 < 1$  and  $0 < \gamma_2 < 1$  and  $h_1 = \gamma_0(1 - \gamma_1 - \gamma_2)^{-1}$ . Lags of  $\varepsilon_t^*$  and  $h_t$  can also be included in the specification (4.11). As a result, the variance matrix of  $\varepsilon_t$  has become time-varying and is given by

$$\Sigma_\varepsilon(h_t) = h_t \Gamma \Gamma' + \Sigma_\varepsilon^+, \quad t = 1, \dots, T, \quad (4.12)$$

where  $\Sigma_\varepsilon(h_t)$  is a full variance matrix but its time-variation depends on the common and univariate GARCH process (4.11). In this specification, the normalization constraint  $\Gamma\Gamma = 1$  can be replaced by fixing  $\gamma_0$  to a known constant. The (unconditional) time-varying variance matrix of  $y_t$  is  $\Lambda(\lambda)\Sigma_\beta\Lambda(\lambda)' + \Sigma_\varepsilon(h_t)$  where  $\Sigma_\beta$  is the solution of  $\Sigma_\beta - \Phi\Sigma_\beta\Phi = \Sigma_\eta$ . The unknown coefficients for the GARCH specification are collected in the parameter vector  $\gamma = (\gamma_1, \gamma_2, \Gamma)'$ . We treat  $\gamma_0$  as a known constant and refer to this model as DNS–GARCH.

In the same framework, we can consider a GARCH specification for  $\Sigma_\eta$  in (4.4) based on the decomposition  $\eta_t = \Gamma_\eta \eta_t^* + \eta_t^+$  with  $\eta_t^* \sim NID(0, g_t)$  and  $\Sigma_\eta = \Sigma_\eta(g_t)$  as in (4.12). When we focus on this decomposition only, the variance matrix of  $y_t$  is given by  $\Lambda(\lambda)\Sigma_\beta(g_t)\Lambda(\lambda)' + \Sigma_\varepsilon$ . In this setting, the variance structure of  $y_t$  subject to volatility depends on matrix  $\Lambda(\lambda)$ . We regard this specification as a restriction compared to (4.12) where the volatility variance structure is determined by  $\Gamma$ . We can construct a likelihood-ratio statistic for this restriction. Another hypothesis can be formulated by considering  $y_t = \Lambda(\lambda)\beta_t + \Gamma\varepsilon_t^* + \varepsilon_t^+$  without GARCH for  $\Sigma_\eta$ . The hypothesis of interest is  $\Gamma = \Lambda(\lambda)w$  where  $w$  is a  $3 \times 1$  vector of unknown coefficients. The null model is then given by  $y_t = \Lambda(\lambda)(\beta_t + w\varepsilon_t^*) + \varepsilon_t^+$ . In this case, the variance matrix of  $y_t$  has become  $\Lambda(\lambda)[\Sigma_\beta + h_t w w']\Lambda(\lambda)' + \Sigma_\varepsilon$  where the variance structure subject to volatility also depends on the factor loadings in  $\Lambda(\lambda)$ .

### 4.3.3 Estimation Based on the Extended Kalman Filter

In case  $\lambda_t$  and  $h_t$  are specified as cubic spline functions depending on a set of knot positions and knot values in  $\lambda^*$  and  $\kappa^*$ , respectively, the Kalman filter methods of Section 4.2.3 for estimation can be used when  $\lambda^*$  and  $\kappa^*$  are placed in the parameter vector  $\psi$ . Although the system matrices  $\Lambda(\lambda_t)$  and  $\Sigma_\varepsilon(h_t)$  of the state space model are time-varying, the Kalman filter is still applicable when these matrices are known a-priori for a given value of  $\psi$ .

In case loading parameter  $\lambda_t$  and variance  $h_t$  (or  $g_t$ ) are specified as an autoregressive process and a GARCH process, respectively, we cannot determine  $\lambda_t$  and  $h_t$  a-priori. In particular,  $\lambda_t$  depends on the Nelson-Siegel latent factors level, slope and curvature ( $\beta_{1,t}$ ,  $\beta_{2,t}$  and  $\beta_{3,t}$ , respectively) while the time-varying variance  $h_t$  in (4.11) depends deterministically on past values of the unobserved disturbance term  $\varepsilon_t^*$ . Therefore, we treat  $\lambda_t = \beta_{4,t}$  and  $\varepsilon_t^*$  as latent variables of interest and place them in the state vector  $\alpha_t$ . A nonlinear state space model can be designed for the DNS model with time-varying parameters based on this state vector. The nonlinear observation equation is given by

$$y_t = Z_t(\alpha_t) + \varepsilon_t^+, \quad \varepsilon_t^+ \sim NID\{0, \Sigma_\varepsilon^+\}, \quad t = 1, \dots, T, \quad (4.13)$$

with  $\alpha_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t}, \varepsilon_t^*)' = (\beta_t', \varepsilon_t^*)'$  and where  $Z_t(\alpha_t)$  is the  $N \times 1$  vector function

$$Z_t(\alpha_t) = \Lambda(\lambda_t)(\beta_{1,t}, \beta_{2,t}, \beta_{3,t})' + \Gamma \varepsilon_t^*, \quad \text{with } \lambda_t = \beta_{4,t}, \quad t = 1, \dots, T. \quad (4.14)$$

The state equation is given by

$$\alpha_{t+1} = c + \begin{bmatrix} \Phi & 0 \\ 0 & 0 \end{bmatrix} \alpha_t + \begin{pmatrix} \eta_t \\ \varepsilon_{t+1}^* \end{pmatrix}, \quad \begin{pmatrix} \eta_t \\ \varepsilon_{t+1}^* \end{pmatrix} \sim NID \left( 0, \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & h_{t+1} \end{bmatrix} \right), \quad (4.15)$$

for  $t = 1, \dots, T$  and  $c = [\mu'(I - \Phi)', 0]'$ . Since  $h_{t+1}$  in (4.11) is a function of the unobserved value  $\varepsilon_t^*$  and its past values, we will not be able to compute the necessary value of  $h_{t+1}$  at time  $t$ . A solution is to replace  $h_{t+1}$  by its estimate based on observations  $y_1, \dots, y_t$ , that is

$$\hat{h}_{t+1|t} = \gamma_0 + \gamma_1 e_t^2 + \gamma_2 \hat{h}_{t|t-1}, \quad t = 1, \dots, T,$$

where  $e_t$  is an estimate of  $\varepsilon_t^*$  based on  $y_1, \dots, y_t$  and obtained from the filtering step of the Kalman filter applied to the model (4.13) and (4.15). Past values of  $\hat{h}_{t|t-1}$  can be stored outside the model and the variance  $h_{t+1}$  in (4.15) is replaced by  $\hat{h}_{t+1|t}$  for the prediction step of the Kalman filter. As a result, the state estimates are sub-optimal, they are not MMSLE. A more detailed discussion of this approach is provided by Harvey, Ruiz, and Sentana (1992).

The Kalman filter method only applies to models that are linear in the state vector. The observation equation (4.13) is clearly nonlinear in  $\alpha_t$ . Exact estimation procedures for nonlinear models require a major computational effort. We prefer to preserve the elegance of the Kalman filter. For this purpose, we locally linearize the nonlinear function  $Z_t(\alpha_t)$  at  $\alpha_t = a_{t|t-1}$  where  $a_{t|t-1}$  is an estimate of  $\alpha_t$  based on the past observations  $y_1, \dots, y_{t-1}$ . We obtain the linearized model

$$y_t = Z_t(a_{t|t-1}) + \dot{Z}_t \cdot (\alpha_t - a_{t|t-1}) + \varepsilon_t^+ = d_t + \dot{Z}_t \alpha_t + \varepsilon_t^+, \quad t = 1, \dots, T,$$

where  $d_t = Z_t(a_{t|t-1}) - \dot{Z}_t a_{t|t-1}$  and  $\dot{Z}_t = \partial Z_t(\alpha_t) / \partial \alpha_t' |_{\alpha_t = a_{t|t-1}} = (\dot{z}'_{1t}, \dots, \dot{z}'_{Nt})'$  with

$$\dot{z}_{it} = \left[ 1, \Lambda_{i2}(a_{4,t|t-1}), \Lambda_{i3}(a_{4,t|t-1}), a_{2,t|t-1} \dot{\Lambda}_{i2}(a_{4,t|t-1}) + a_{3,t|t-1} \dot{\Lambda}_{i3}(a_{4,t|t-1}), \Gamma_i \right],$$

for which the loading element  $\Lambda_{ij}(\lambda)$  is given below (4.3),  $\dot{\Lambda}_{ij}(x) = \partial \Lambda_{ij}(\lambda) / \partial \lambda |_{\lambda=x}$ ,  $a_{k,t|t-1}$  is the  $k$ th element of vector  $a_{t|t-1}$  and  $\Gamma_i$  is the  $i$ th element of vector  $\Gamma$ , for  $i = 1, \dots, N$ ,  $j = 2, 3$  and  $k = 2, 3, 4$ . Given an estimate  $a_{t|t-1}$  and an approximate MSE matrix  $A_{t|t-1}$  for  $a_{t|t-1}$ , the filtering step is given by

$$a_{t|t} = a_{t|t-1} + A_{t|t-1} \dot{Z}_t' F_t^{-1} v_t, \quad A_{t|t} = A_{t|t-1} - A_{t|t-1} \dot{Z}_t' F_t^{-1} \dot{Z}_t A_{t|t-1}, \quad (4.16)$$

with  $v_t = y_t - d_t - \dot{Z}_t a_{t|t-1} = y_t - Z_t(a_{t|t-1})$  and  $F_t = \dot{Z}_t A_{t|t-1} \dot{Z}_t' + \Sigma_\varepsilon^+$ . We define  $a_{t|t}$  as a sub-optimal estimate of  $\alpha_t$  based on observations  $y_1, \dots, y_t$  and  $A_{t|t}$  as its approximate MSE matrix. The prediction step is similar to (4.6) but then based on the state equation (4.15).

The estimates  $a_{t|t-1}$  and  $a_{t|t}$  are sub-optimal due to the replacement of  $h_{t+1}$  in (4.15) by  $\hat{h}_{t+1}$  and due to the linearization of the original observation equation (4.13). We therefore label  $A_{t|t-1}$  and  $A_{t|t}$  as approximate MSE matrices. For a given time series  $y_1, \dots, y_T$ , the filtering and prediction steps can be carried out recursively for  $t = 1, \dots, T$ . The resulting algorithm is known as the extended Kalman filter, see Anderson and Moore (1979) for a more formal derivation. The quasi-loglikelihood function is obtained by inserting the values  $v_t$  and  $F_t$ , defined below (4.16), into the loglikelihood (4.7). We then maximize the quasi-likelihood to obtain estimates for  $\psi$ . Estimates of the latent Nelson-Siegel factors, the loading parameter  $\lambda_t$  and the GARCH variance  $h_t$  (or  $g_t$ ) are based on the filtered state estimate  $a_{t|t}$ .

## 4.4 Data and Empirical Findings

For our empirical analysis of yield curves we consider the unsmoothed Fama-Bliss zero-coupon yields dataset, obtained from the CRSP unsmoothed Fama and Bliss (1987) forward rates. We analyze monthly U.S. Treasury yields with maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months over the period from January 1972 to December 2000. This dataset is the same as the one analyzed by Diebold, Rudebusch, and Aruoba (2006) and Diebold and Li (2006) who provide more details on its construction.

**Table 4.1: Summary Statistics**

The table reports summary statistics for U.S. Treasury yields over the period 1972-2000. We examine monthly data, constructed using the unsmoothed Fama-Bliss method. Maturity is measured in months. For each maturity we show mean, standard deviation (*St Dev*), minimum, maximum and three autocorrelation coefficients, 1 month ( $\hat{\rho}(1)$ ), 1 year ( $\hat{\rho}(12)$ ) and 30 months ( $\hat{\rho}(30)$ ).

Summary Statistics for each Maturity							
Maturity	Mean	St Dev	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	6.851	2.695	2.732	16.020	0.970	0.700	0.319
6	7.079	2.702	2.891	16.481	0.972	0.719	0.355
9	7.201	2.679	2.984	16.394	0.972	0.726	0.378
12	7.302	2.602	3.107	15.822	0.971	0.729	0.394
15	7.408	2.548	3.288	16.043	0.973	0.737	0.415
18	7.481	2.532	3.482	16.229	0.974	0.743	0.431
21	7.544	2.520	3.638	16.177	0.975	0.747	0.442
24	7.558	2.474	3.777	15.650	0.975	0.745	0.450
30	7.647	2.397	4.043	15.397	0.975	0.755	0.470
36	7.724	2.375	4.204	15.765	0.977	0.761	0.480
48	7.861	2.316	4.308	15.821	0.977	0.765	0.499
60	7.933	2.282	4.347	15.005	0.980	0.779	0.514
72	8.047	2.259	4.384	14.979	0.980	0.786	0.524
84	8.079	2.215	4.352	14.975	0.980	0.768	0.526
96	8.142	2.201	4.433	14.936	0.982	0.793	0.535
108	8.176	2.209	4.429	15.018	0.982	0.794	0.540
120(level)	8.143	2.164	4.443	14.925	0.982	0.771	0.532
slope	1.292	1.461	-3.505	4.060	0.929	0.410	-0.099
curvature	0.121	0.720	-1.837	3.169	0.788	0.259	0.076

Table 4.1 provides summary statistics for our dataset. For each maturity, we report mean, standard deviation, minimum, maximum and some autocorrelation coefficients. We also present the statistics for proxies of the level, slope and curvature of the yield curve, see the discussion in Section 4.2.1. The summary statistics reveal that the average



yield curve is upward sloping. Volatility decreases by maturity, with the exception of the 6-month being more volatile than the 3-month bill. Yields for all maturities are persistent, most notably for long term bonds. However, with a first-order autocorrelation of 0.970, the 3-month bill is also highly persistent. The level, slope and curvature proxies are persistent but to a lesser extent. The curvature and slope proxies are least persistent given the twelfth-order autocorrelation coefficients of 0.259 and 0.410, respectively.

**Figure 4.1: Yield Curves from January 1972 up to December 2000**

In this figure the U.S. Treasury yields over the period 1972-2000 are shown. We examine monthly data, constructed using the unsmoothed Fama-Bliss method. The maturities we show are 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months.

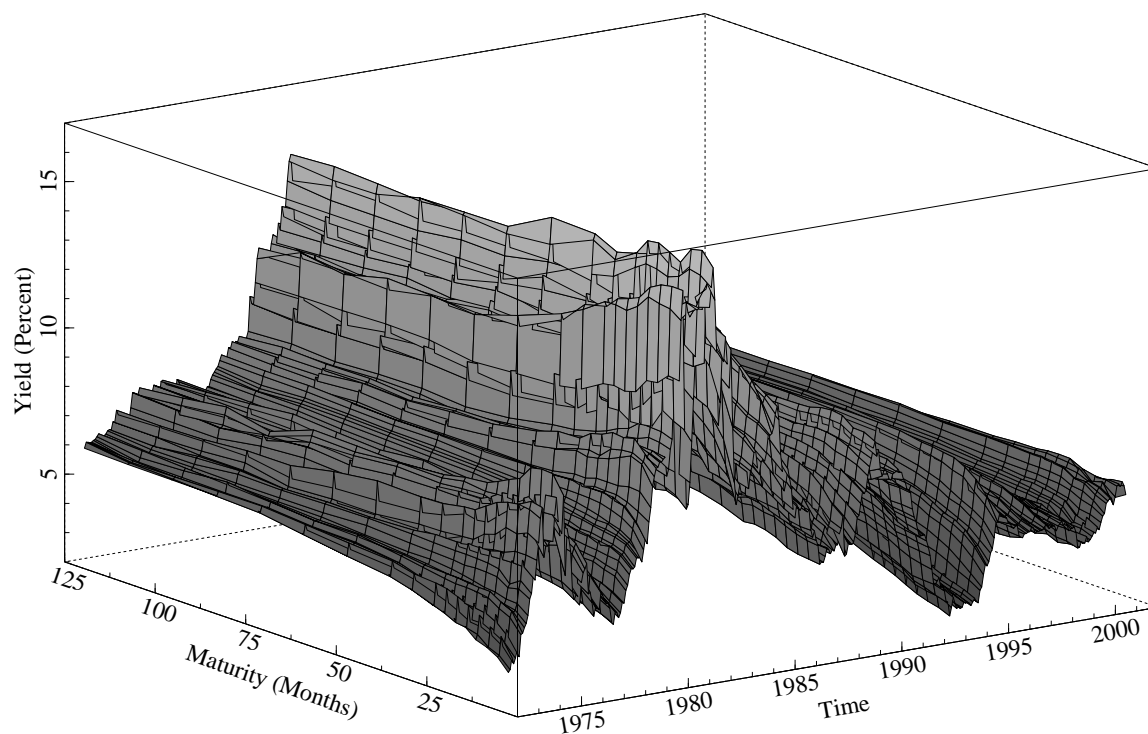


Figure 4.1 shows the cross-section of yields over time. This is a graphical representation of our data-set. In addition to the findings of Table 4.1 we see a few interesting characteristics. The first noticeable fact is that yields vary significantly over time from which various common dynamics across all yields can be deduced. Especially in the years 1978-1987 interest rates are remarkable high and volatile. Second, the shape of the yield curve is not constant over time. Though on average it is upward sloping, there are periods when it is downward sloping or humped.

### 4.4.1 DNS: Baseline Dynamic Nelson-Siegel Model

**Table 4.2: Baseline Model - Estimates of VAR Model for Latent Factors**

The table reports the estimates of the vector autoregressive (VAR) model for the latent factors. The results shown correspond to the latent factors of the baseline Nelson-Siegel latent factor model (DNS). Panel A shows the estimates for the constant vector  $\mu$  and autoregressive coefficient matrix  $\Phi$ , Panel B shows the variance matrix  $\Sigma_\eta$ .

Panel A: Baseline DNS Model				
Constant and Autoregressive Coefficients of VAR				
	Level <sub>t-1</sub> ( $\beta_{1,t-1}$ )	Slope <sub>t-1</sub> ( $\beta_{2,t-1}$ )	Curvature <sub>t-1</sub> ( $\beta_{3,t-1}$ )	Constant ( $\mu$ )
Level <sub>t</sub> ( $\beta_{1,t}$ )	0.997** 0.00811	0.0271** 0.00889	-0.0216* 0.0105	8.03** 1.27
Slope <sub>t</sub> ( $\beta_{2,t}$ )	-0.0236 0.0167	0.942** 0.0176	0.0392 0.0212	-1.46** 0.527
Curvature <sub>t</sub> ( $\beta_{3,t}$ )	0.0255 0.023	0.0241 0.0257	0.847** 0.0312	-0.425 0.537

*\*/\*\* indicates significance at the 95%/99% level. The standard errors are reported below the estimates.*

Panel B: Baseline Model - Variance Matrix of VAR			
	Level <sub>t</sub> ( $\beta_{1,t}$ )	Slope <sub>t</sub> ( $\beta_{2,t}$ )	Curvature <sub>t</sub> ( $\beta_{3,t}$ )
Level <sub>t</sub> ( $\beta_{1,t}$ )	0.0949** 0.00841	-0.014 0.0113	0.0439* 0.0186
Slope <sub>t</sub> ( $\beta_{2,t}$ )		0.384** 0.0306	0.00927 0.0344
Curvature <sub>t</sub> ( $\beta_{3,t}$ )			0.801** 0.0812

*\*/\*\* indicates significance at the 95%/99% level. The standard errors are reported below the estimates.*

Table 4.2 presents the estimates of the vector autoregression (VAR) model for the latent factors. The high persistence from the proxies for the level, slope and curvature that we report in Table 4.1 are confirmed by the high diagonal elements of the VAR coefficient matrix. The estimates in this table are almost identical to those in Diebold, Rudebusch, and Aruoba (2006, Table 1, p.316). A slight difference stems from our restriction of a stationary VAR process for the factors, see Section 4.2.2. The factor loadings parameter  $\lambda$  is estimated as 0.0778, with a standard error of 0.00209. The high significance of this estimate confirms that interest rates are informative about  $\lambda$  while small changes in the loadings have a significant effect on the likelihood value.

Table 4.3 reports sample means and standard deviations of filtered errors. The filtered errors are defined as the difference between the observed yield curve and its filtered estimate, obtained from the Kalman filter. We find that in particular the 3-month rate is difficult to fit: it has the highest mean filtered error. The standard deviations reported in

**Table 4.3: Filtered Errors of Model Extensions**

The table reports the filtered errors from the four Nelson-Siegel latent factor models we estimate. The filtered errors are defined as the difference between the observed yield curve and its filtered estimate, obtained from the Kalman filter. The *Baseline* model corresponds to the baseline dynamic Nelson-Siegel latent factor model with constant factor loadings and volatility (DNS). The *Time-Varying Factor Loading* model corresponds to the model with  $\lambda$  added to the state (DNS–TVL). The *Time-Varying Volatility* model corresponds to the model with a common GARCH component for the volatility (DNS–GARCH). The *Both Time-Varying* model corresponds to the model with the factor loadings parameter added to the state and the common GARCH component for volatility (DNS–TVL–GARCH). For each maturity we show mean and standard deviation (*St Dev*). We summarize these per model with three statistics: the mean, median and number of maturities for which the absolute value is lower than that of the baseline model (*#Lower*).

Filtered Errors (in basis points)								
Maturity	Baseline Model DNS		Time-Varying Factor Loading DNS–TVL		Time-Varying Volatility DNS–GARCH		Both Time-Varying DNS–TVL –GARCH	
	Mean	St Dev	Mean	St Dev	Mean	St Dev	Mean	St Dev
3	-12.63	22.37	-2.87	14.15	-7.61	21.88	-3.29	16.02
6	-1.34	4.87	0.19	1.99	0	0	0.17	0.92
9	0.51	8.13	-0.95	7.54	0.01	9.73	-0.77	7.9
12	1.32	9.89	-0.89	9.46	-0.6	10.53	-1.14	9.61
15	3.72	8.76	1.71	8.29	0.16	5.83	-0.06	5.84
18	3.63	7.22	2.15	6.38	0.72	4.18	0.63	3.73
21	3.26	6.43	2.39	5.82	1.45	6.15	1.87	6.19
24	-1.39	6.33	-1.69	7	-0.87	7.38	-0.66	6.71
30	-2.68	5.98	-2.11	6.35	-1.5	6.42	-1.41	6.05
36	-3.29	6.6	-2.22	6.71	-1.21	5.71	-1.37	5.74
48	-1.83	9.67	-0.52	9.19	1.11	7.6	0.78	7.86
60	-3.29	7.98	-2.3	7.15	-0.79	5.58	-1.46	6.84
72	1.94	9.02	2.41	8.68	2.61	9.01	2.18	9
84	0.68	10.18	0.59	10.6	0.21	11.05	0.58	10.84
96	3.51	9.15	2.9	9.9	1.84	8.83	1.66	8.75
108	4.24	13.5	3.16	13.22	-0.54	7.81	0.48	8.89
120	-1.33	16.34	-2.82	16.43	-4.09	14.7	-3.17	15.82
Mean	-0.29	9.55	-0.05	8.76	-0.54	8.38	-0.29	8.04
Median	0.51	8.76	-0.52	8.29	0	7.6	-0.06	7.86
#Lower			13	11	15	12	14	14

Table 4.3 indicate that the bonds with intermediate maturity are filtered most accurately.

#### 4.4.2 DNS–TVL: Time-Varying Factor Loadings

To obtain some indication whether the  $\lambda$  parameter varies over time, we consider the baseline model for four equally sized subperiods that cover the full sample. The four estimates of  $\lambda$  for the consecutive subperiods are 0.0397, 0.126, 0.0602 and 0.0695. The corresponding standard errors are sufficiently small to conclude that the four  $\lambda$  estimates are distinct from each other (except for the last two subsamples). This finding provides some evidence that the assumption of constant factor loadings over time does not necessarily hold.

Further evidence is obtained by treating the factor loadings parameter  $\lambda$  as a time-varying spline function, see Section 4.3.1. We use a spline function based on five knots which are equally spaced over the sample. Panel (A) of Figure 4.2 presents the estimate of the spline function for  $\lambda$ . It is evident from the graph that the factor loadings parameter  $\lambda$  is not constant over time.

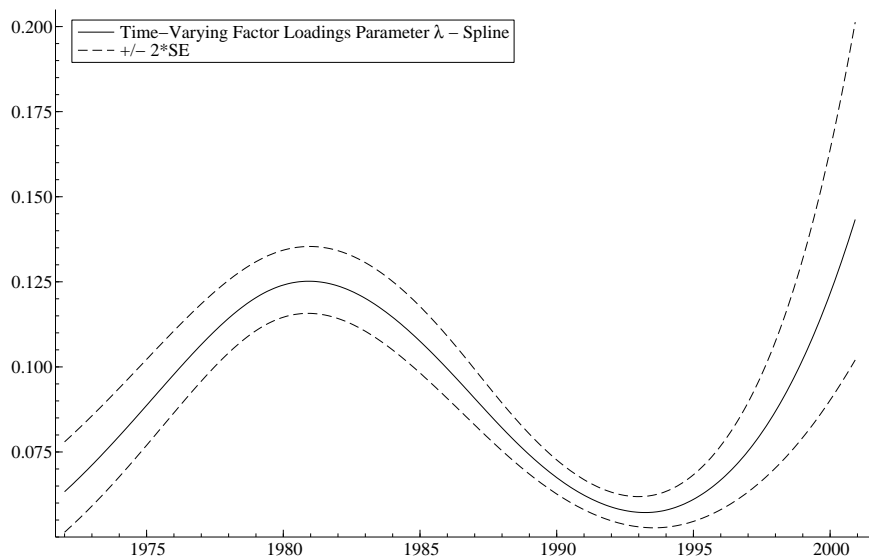
Different number of knots and knot positions will oftentimes lead to different spline functions and possibly different conclusions with respect to the time-varying nature of the yield curve. We therefore consider an alternative to the splines approach and modify the model specification by treating the factor loadings parameter  $\lambda$  as a latent factor that is modeled jointly with the other factors by a VAR process, see Section 4.3.1. We estimate the coefficients of this model and obtain filtered estimates of both the three yield factors and the time-varying  $\lambda$  using the extended Kalman filter discussed in Section 4.3.3. Panel (A) of Figure 4.3 presents the filtered estimates of the factor loadings parameter  $\lambda$ . The  $\lambda$  estimates in 1974 are particularly high whereas at the end of the 1970's and the beginning of the 1980's the estimates are rather volatile. Although many changes occur in the early part of the sample, the changes in the late 1990s are also pronounced. Since both slope and curvature of the yield depend on  $\lambda$ , we conclude that sufficient evidence is provided of significant changes in the characteristics of the yield curve over time.

Parameter estimates of the DNS model with  $\lambda$  as a latent factor are discussed in Section 4.4.4. Here we focus on the fit of the model. Table 4.3 enables comparisons, for each maturity, between the sample means of the filtered errors for the DNS and DNS–TVL models. For 13 out of the 17 maturities the mean filtered error is lower. This is particularly the case for short maturities. The standard deviations of the filtered errors are lower for 11 out of the 17 maturities. Table 4.4 reports the performance of the DNS models by presenting values for the loglikelihood, the Akaike Information Criterion (AIC) together with the likelihood-ratio (LR) value. When comparing the loglikelihood values between the DNS and DNS-TVl models, the difference of 300 is convincing by any means. This is confirmed by the AIC and LR values. The results therefore provide sufficient evidence

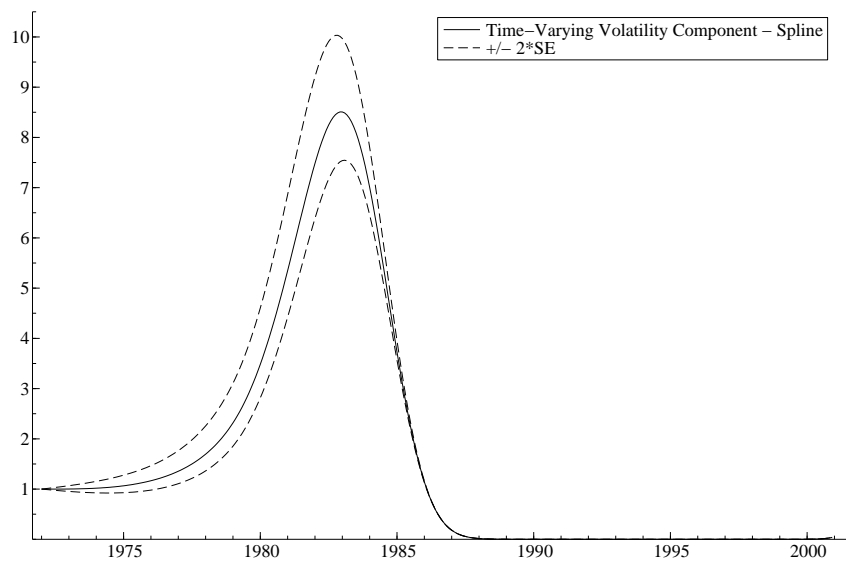
### Figure 4.2: Time-Varying Factor Loadings and Volatility using Splines

In this figure we present the time-varying factor loadings parameter and volatility modeled with splines. The knots for the splines are chosen at both the beginning and end of the sample, and at April 1979, July 1986 and October 1993.

(A) Time-Varying Factor Loadings, modeled using Spline



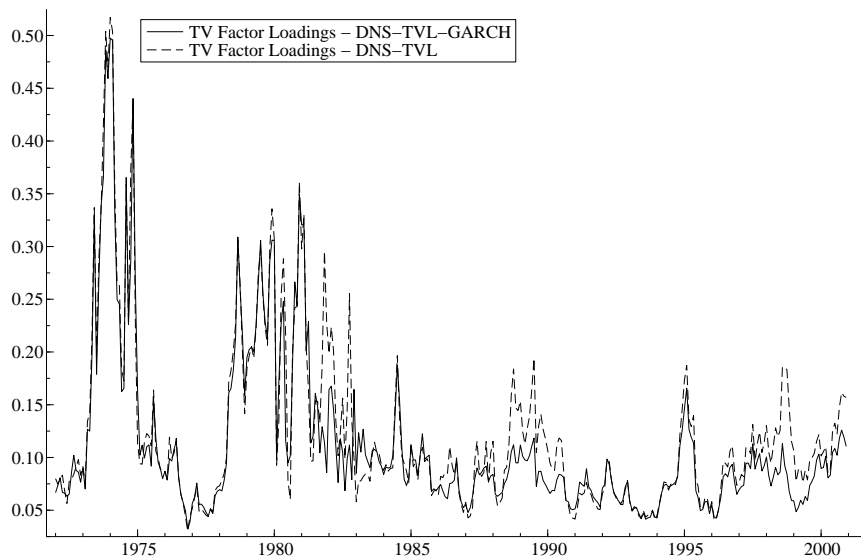
(B) Common Time-Varying Volatility Component, modeled using Spline



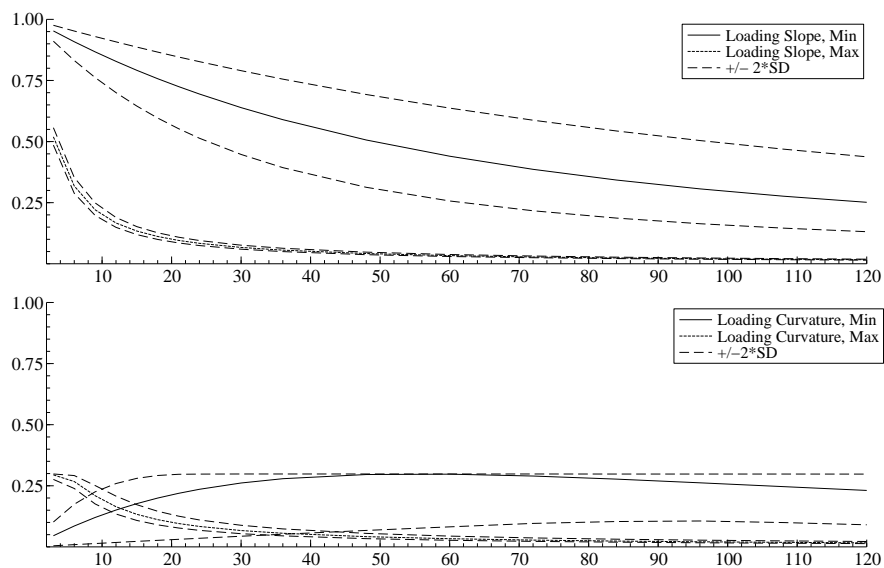
**Figure 4.3: Time-Varying Factor Loadings Parameter Added to State**

In this figure we present the filtered time series of the factor loadings parameter  $\lambda$  and the slope and curvature loadings using the minimum and maximum value of the filtered  $\lambda$ . In Panel (A) we show the filtered time series for both the model with the factor loadings parameter added to the state (DNS-TVL, dashed line) and the model with both the time-varying factor loadings parameter added to the state and a common GARCH component for the volatility (DNS-TVL-GARCH, solid line). In Panel (B) we show the slope and curvature loadings using the minimum and maximum value of the filtered  $\lambda$  for the DNS-TVL-GARCH model.

(A) Filtered Factor Loadings Parameter



(B) Minimum and Maximum loading of Slope and Curvature



**Table 4.4: Loglikelihood and AIC of Model Extensions**

Panel A reports the loglikelihood and Akaike Information Criterion (AIC) for the various model extensions proposed. The *Baseline* model corresponds to the baseline dynamic Nelson-Siegel latent factor model with constant factor loadings and volatility (DNS). The *Time-Varying Factor Loadings* model corresponds to the model with  $\lambda$  added to the state (DNS–TVL). The *Time-Varying Volatility* model corresponds to the model with a common GARCH component for the volatility (DNS–GARCH). The *Time-Varying Loadings and Volatility* model corresponds to the model with the factor loadings parameter added to the state and the common GARCH component for volatility (DNS–TVL–GARCH). In Panel B we report the loglikelihood and AIC for various alternative models and for our extensions estimated only for the period after 1987.

Panel A: Performance of Model Extensions				
	Loglikelihood	AIC	#Parameters	LR-value
DNS–baseline	3184.6	-6297.1	36	
DNS–TVL	3484.9	-6875.7	47	600.6
DNS–GARCH	3657.3	-7204.7	55	945.6
DNS–TVL–GARCH	3766.8	-7401.7	66	1,164.6

Panel B: Alternative Models and Results for Post-1987 Period				
	Loglikelihood	AIC	#Parameters	LR-value
<i>Alternative Model Specifications</i>				
DNS–GARCH (in $\eta_t$ )	3199.1	-6316.2	41	29.1
DNS–GARCH ( $\Gamma = \Lambda(\lambda)w$ )	3276.6	-6471.2	41	184.1
<i>Models Estimated for Post-1987 Period</i>				
DNS–baseline (>1987)	3041.7	-6011.4	36	
DNS–TVL (>1987)	3213.5	-6333.0	47	343.6
DNS–GARCH (>1987)	3544.3	-6978.6	55	1,005.2
DNS–TVL–GARCH (>1987)	3668.5	-7205.0	66	1,253.6

of a highly significant improvement in the fit of the DNS–TVL model over its baseline version.

### 4.4.3 DNS–GARCH: Time-Varying Volatility

The second modification we propose in Section 4.3 is to allow for time-varying volatility. First, we introduce a spline function for the time-varying common volatility pattern in the observation disturbances. The position of the knots are set equal to those of Section 4.4.2 for  $\lambda$ . Panel (B) of Figure 4.2 presents the estimated spline for the common volatility. It shows that volatility is high in the period between 1980 and 1987. Thereafter it is almost completely constant for all maturities.

The use of a spline function for time-varying volatility can be subjected to the same criticism as for its use of a time-varying  $\lambda$ . Therefore we allow for a common time-varying volatility component in the observation disturbances using the GARCH specification discussed in Section 4.3.2. The details of estimation are discussed in Section 4.3.3. Panel (A) of Figure 4.4 presents the filtered estimates of the common volatility. It shows that the common volatility is particularly high in the early years of the 1980’s while from the end of the 1980’s onwards the volatility is low and rather constant over time. The latter finding may suggest that after the publication of Nelson and Siegel (1987), their method has become the default of practitioners to price the cross-section of yields which may have had a dampening effect on volatility. However, low volatility in a prolonged period from the mid-1980’s has also been detected for time series of U.S. Inflation, see the discussion in Stock and Watson (2007).

Table 4.3 reports the mean of the filtered errors for the model with GARCH and this mean is lower for 15 out of the 17 maturities when compared to those for the baseline DNS model. Only the 72 and 120-month bonds have a higher mean in the DNS–GARCH model. Furthermore, the standard deviations of the filtered errors of the DNS–GARCH model is lower for 12 out of the 17 maturities. In Table 4.4 we compare loglikelihood and AIC values of the DNS–GARCH model with those of the baseline DNS model. Similarly to the DNS–TVL model, we find a highly significant improvement in the loglikelihood value of the DNS–GARCH model over the baseline model. The likelihood increase and the AIC decrease are even higher than in the case for the DNS–TVL models. It indicates that most gains in describing the yield curve in this dataset are obtained by introducing time-varying volatility.

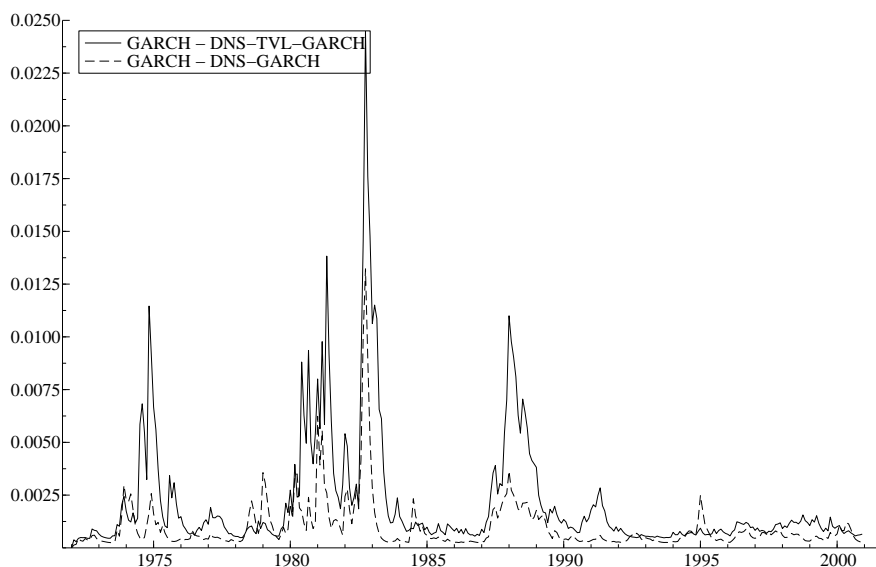
We also consider the DNS–GARCH model for treating volatility in  $\eta_t$ , the innovations of the factors in (4.4). In this specification, the GARCH process is loaded onto the level, slope and curvature factors while it is indirectly loaded onto the observed yields via  $\beta_t$ . Empirical support for this specification is weak, the GARCH parameter estimates



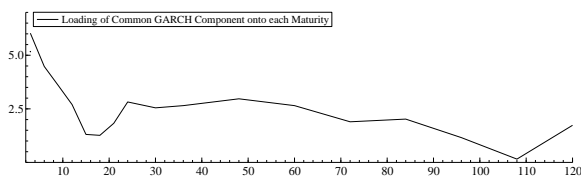
**Figure 4.4: Time-Varying Volatility**

In this figure we present the time-varying volatility. In Panel (A) we plot the time varying volatility for both the model with a common GARCH volatility component (DNS-GARCH, dashed line) and the model with both time-varying factor loadings and a common GARCH component for the volatility (DNS-TVL-GARCH, solid line). In Panel (B) we show the loadings, for each maturity, of the common GARCH process in the DNS-TVL-GARCH model. Panel (C) shows the estimated volatility for the DNS-TVL-GARCH model.

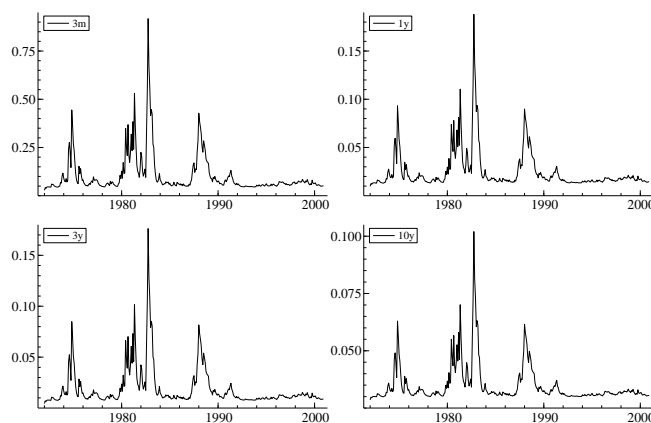
(A) Filtered Common GARCH Volatility Component



(B) Loadings of Common GARCH Volatility Component against Maturity



(C) Estimated Volatility for Some Maturities



indicate that the common volatility component is close to a constant while the other parameter estimates are similar to those obtained for the DNS model. The loglikelihood increase of 14.5 reported in Panel B of Table 4.4 is relatively small but it is significant. However, we obtain stronger support for a common GARCH component in the observation disturbances in  $\varepsilon_t$ . In the latter case, we can examine whether the GARCH loadings are linear combinations of the factor loadings via the restriction  $\Gamma = \Lambda(\lambda)w$  where  $w$  is unknown and needs to be estimated. The resulting loglikelihood increase of 92 compared to the baseline model is significant but moderate when compared to the increase obtained by the unrestricted DNS–GARCH model.

#### 4.4.4 DNS–TVL–GARCH: Time-Varying Loadings and Volatility

Given the encouraging initial results of the last two subsections, we next discuss in more detail the estimation results for the DNS–TVL–GARCH model, the DNS model with both time-varying factor loadings and volatility. In Table 4.3, the means and standard deviations of the filtered errors for the full model specification are given. In comparison with the baseline DNS model, we observe that the filtered error mean is lower for 14 out of 17 maturities. Although this improvement is slightly less than for the DNS–GARCH model, we also have 14 error series that have smaller standard deviations compared to the baseline model. Such improvement has not been obtained by the other DNS models.

The loglikelihood and AIC values reported in Panel A of Table 4.4 for the full model show strong significant improvements compared to the baseline DNS model. When we benchmark the values against models with only time-varying factor loadings or only time-varying volatility, we also obtain significant improvements. We therefore conclude that both model extensions significantly contribute to improvements in the DNS model fit. The GARCH extension provides the most significant improvement.

Panel (A) of Figure 4.3 presents the filtered  $\lambda$  estimates obtained from the DNS–TVL–GARCH model where  $\lambda$  is treated as a latent factor. The  $\lambda$  estimates are similar to the DNS–TVL model. Panel (B) presents the loadings for slope and curvature that are obtained using the minimum and maximum value of the estimates of  $\lambda$ . It shows clearly that the loadings can differ significantly over time. The current model specification provides this flexibility. In Panel (A) of Figure 4.4 the filtered estimate of the common GARCH component is displayed for the DNS–TVL–GARCH model. The volatility estimates are similar to the DNS–GARCH model. However, in the period at the end of the 1980's, the estimates of both  $\lambda$  and the common volatility are different when compared to the single DNS extensions. It is interesting to observe that for this period the filtered  $\lambda$  estimates are lower compared to the DNS–TVL model. The sharp increases in the yields in this pe-

riod are explained more accurately by a common GARCH component than a time-varying loading parameter  $\lambda$ .

**Table 4.5: Estimates of Latent Factors VAR Model and GARCH Process**

The table reports the estimates of the vector autoregressive (VAR) model for the latent factors and the GARCH parameter estimates. The results shown correspond to the latent factors of the Nelson-Siegel latent factor model with the time-varying factor loadings parameter added to the state and a common GARCH component for the volatility (DNS-TVL-GARCH). Panel A shows the estimates for the constant vector  $\mu$  and autoregressive coefficient matrix  $\Phi$ , Panel B shows the estimates for the covariance matrix  $\Sigma_\eta$ , Panel C the estimates for the common GARCH process.

Panel A: Constant and Autoregressive Coefficients of VAR					
	Level $_{t-1}$	Slope $_{t-1}$	Curvature $_{t-1}$	Loading $_{t-1}$	Constant ( $\mu$ )
Level $_t$ ( $\beta_{1,t}$ )	0.994** 0.00832	0.0497** 0.016	-0.0287* 0.0135	0.0369 0.0437	7.82** 1.24
Slope $_t$ ( $\beta_{2,t}$ )	-0.0118 0.0133	0.931** 0.0298	0.0149 0.0255	-0.0165 0.0629	-1.63** 0.409
Curvature $_t$ ( $\beta_{3,t}$ )	-0.0308 0.0338	0.198** 0.0646	0.658** 0.0449	0.878** 0.17	0.443 0.392
Loading $_t$ ( $\lambda_t$ )	0.0179 0.00948	-0.0555** 0.0177	0.0734** 0.0165	0.585** 0.0639	-2.37** 0.143

\*/\*\* indicates significance at the 95%/99% level. The standard errors are reported below the estimates.

Panel B: Variance Matrix of VAR				
	Level $_t$ ( $\beta_{1,t}$ )	Slope $_t$ ( $\beta_{2,t}$ )	Curvature $_t$ ( $\beta_{3,t}$ )	Loading $_t$ ( $\lambda_t$ )
Level $_t$ ( $\beta_{1,t}$ )	0.0988** 0.00899	-0.0445** 0.0116	0.0986** 0.0367	-0.00393 0.0157
Slope $_t$ ( $\beta_{2,t}$ )		0.237** 0.0365	0.0596 0.0537	0.0595** 0.0203
Curvature $_t$ ( $\beta_{3,t}$ )			1.6** 0.22	-0.302** 0.0638
Loading $_t$ ( $\lambda_t$ )				0.188** 0.0265

\*/\*\* indicates significance at the 95%/99% level. The standard errors are reported below the estimates.

Panel C: GARCH Parameters			
	$\gamma_0$	$\gamma_1$	$\gamma_2$
Estimate	0.0001	0.471**	0.506**
	NA	0.118	0.118

\*/\*\* indicates significance at the 95%/99% level.

The standard errors are reported below the estimates.

In Table 4.5 we report a selection of the parameter estimates for the DNS-TVL-GARCH model. We first focus on the estimate of the VAR coefficient matrix  $\Phi$  for  $\beta_t$ , with the four latent factors, which is reported in Panel A of Table 4.5. When compared to the estimates of the baseline DNS model, reported in Table 4.2 and by Diebold, Rudebusch, and Aruoba (2006, Table 1, p.316), the inclusion of  $\lambda$  as a latent factor mostly affects the dynamics of the slope and curvature. A new empirical finding is the high persistence of the time-varying factor loadings parameter  $\lambda$ . The results also reveal that the curvature

factor depends heavily on the factor loadings parameter while, compared to the baseline model, it is less persistent and has a higher variance. The factor loadings parameter  $\lambda$  depends heavily on the (lagged) slope and curvature factors. The estimated variance matrix  $\Sigma_\eta$  is reported in Panel B. Although the four innovation series for the factors are all correlated, the strong negative correlation between the curvature factor and the  $\lambda$  factor suggests a substitution effect.

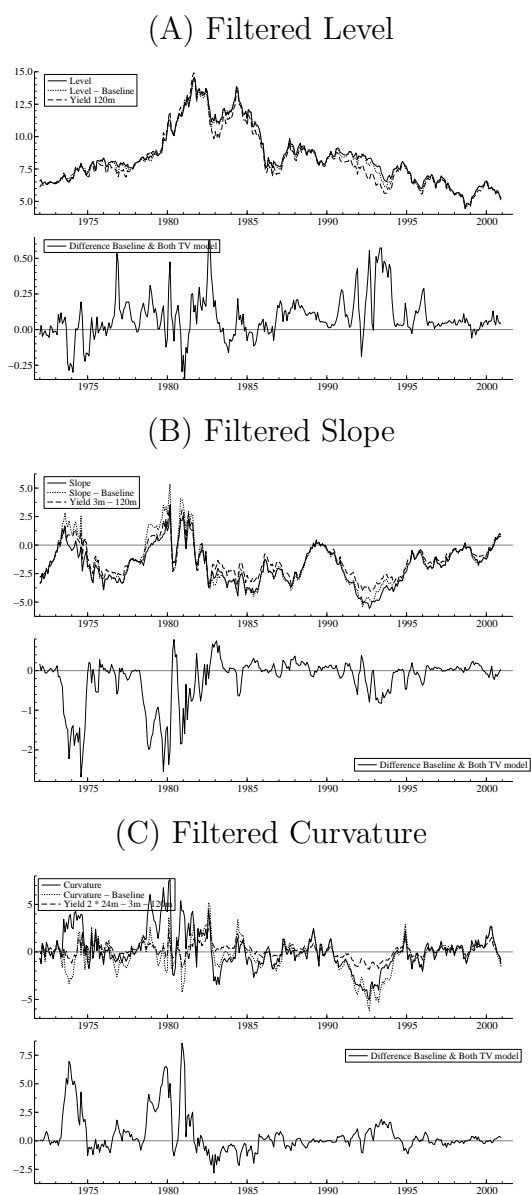
The estimates of the GARCH parameters are presented in Panel C of Table 4.5. Since we estimate all elements in loading vector  $\Gamma$ , the constant in the GARCH specification cannot be identified and is kept at a fixed small value. The remaining estimates for the coefficients  $\gamma_1$  and  $\gamma_2$  are significant and they have similar values as the ones for the DNS–GARCH model (not reported here). The estimates of the elements in  $\Gamma$  are presented graphically in Panel (B) of Figure 4.4. The estimated loadings are displayed by a line-plot against the maturity length. Although the loadings are quite smooth against maturity, it is interesting to find that the maturities of 15 and 18 are relatively less subject to the common GARCH component while the short maturities are most affected by GARCH. When the estimated loadings in  $\Gamma$  are interacted with the GARCH component, we obtain the time-varying volatility for each maturity. Panel (C) in Figure 4.4 displays the volatility process for a selection of maturities.

In Figure 4.5 we compare the filtered latent factors obtained from the DNS–TVL–GARCH model with those from the baseline model and their data-based proxies. The level factors are presented with the 120 month yield, the slopes with the spread of 3 month over 120 month yields and the curvatures with the 24 month yield minus the 3 and 120 month yield. The estimated factors from both models describe the data-based proxies equally well. To highlight the differences in fit of our model extensions, the bottom plots in each panel of Figure 4.5 present the differences of the factors between the DNS and DNS–TVL–GARCH models. The differences are most pronounced for the slope and curvature factors, particularly in the 1973-1974, 1978-1983 and 1991-1994 periods. It confirms the findings reported in Table 4.5 from which we learn that the dynamics for slope and curvature have been most affected by our extensions when compared to the baseline DNS model.

Finally, in Figure 4.6 we discuss the four selected fitted yield curves as earlier reported in Diebold and Li (2006, Figure 5). We report the yield curve obtained from the DL OLS model, DRA SSF model (DNS) and our extended SSF model with both the factor loadings and volatility time-varying (DNS–TVL–GARCH). Especially in August 1998 it is clear that our model extensions allow for more flexibility and improve the fit of the model.

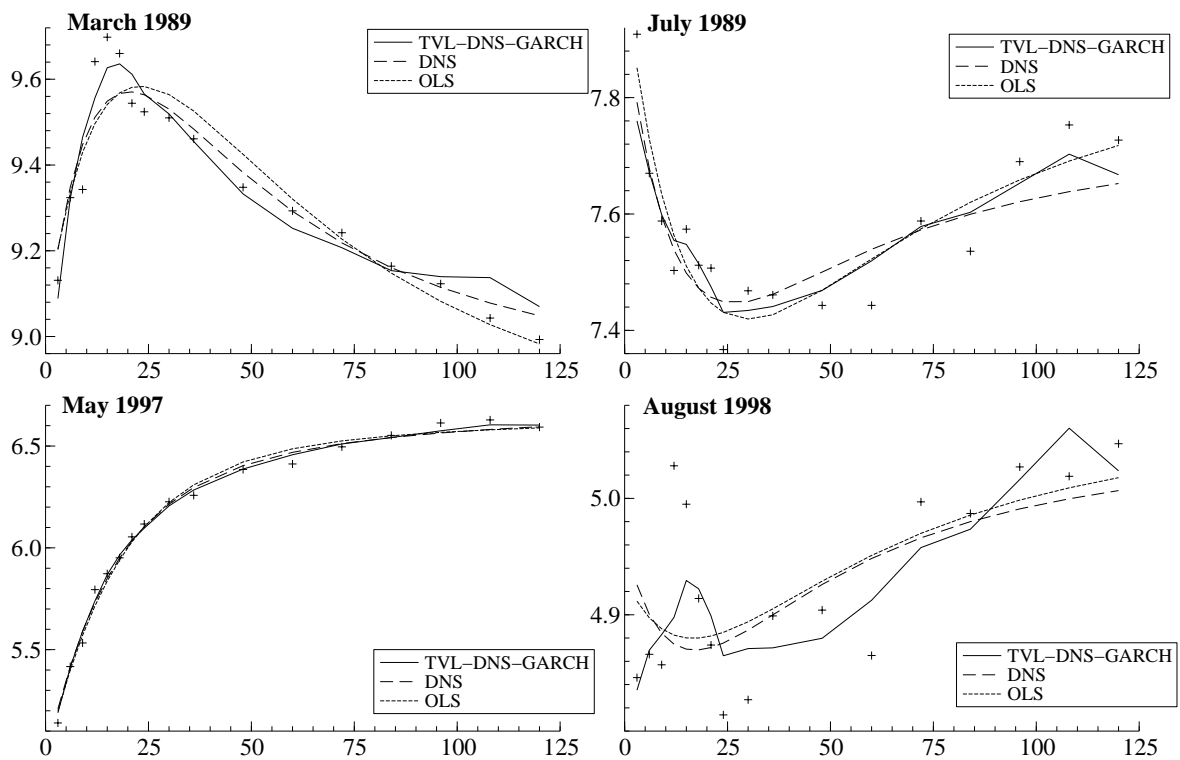
**Figure 4.5: Level, Slope and Curvature**

This figure reports the level, slope and curvature as obtained from the Nelson-Siegel latent factor model with both time-varying factor loadings and volatility (DNS-TVL-GARCH). Panels (A), (B) and (C) report the level, slope and curvature respectively together with their proxies from the data. For the level this is the 120 month treasury yield, for slope this is the spread of 3 month over 120 month yields and for curvature this is twice the 24 month yield minus the 3 and 120 month yield. In addition we show the filtered level, slope and curvature for the baseline dynamic Nelson-Siegel model (DNS) and the difference compared to the latent factors from the DNS-TVL-GARCH model (bottom plots in each panel).



**Figure 4.6: Fitted Yield Curves for Four Months**

This figure presents the fitted yield curve obtained from the Nelson-Siegel latent factor model with both time-varying factor loadings and volatility (DNS-TVL-GARCH). The dots represent the actual yield curve, the solid line the fitted yield curve obtained from the Nelson-Siegel latent factor with both our extensions (DNS-TVL-GARCH), the dashed line the model as put in state space form by Diebold, Rudebusch, and Aruoba (2006, DNS) and the dotted line the OLS model as in Diebold and Li (2006). We show these for four different months: March 1989, July 1989, May 1997 and August 1998.



### 4.4.5 Robustness of Empirical Results

In this section we study the robustness of our results in three ways: (a) comparison with regression results; (b) model with time-varying splines; (c) results based on a different sample.

**(a) Results based on regression.** When the VAR specification is discarded in the DNS model, the original Nelson and Siegel (1987) model (4.2) is obtained and the factors level  $\beta_{1t}$ , slope  $\beta_{2t}$  and curvature  $\beta_{3t}$  can be estimated for each period  $t$  using standard regression methods. In case  $\lambda$  is treated as unknown, it can be estimated by nonlinear least squares (NLS), see Diebold and Li (2006). For a given estimate of  $\lambda$ , the factors and the constant variance  $\sigma_t^2$  in (4.2) can be estimated by ordinary least squares (OLS).

In Panel (A) of Figure 4.7, the NLS estimates of  $\lambda$  in the Nelson-Siegel model are displayed (as dots) together with the estimates of factor  $\beta_{4t}$  in the DNS–TVL–GARCH model (solid line) as obtained from the methods described in Section 4.3.3. The individual NLS estimates are well represented by the estimated fourth factor. In some cases, the  $\lambda$  parameter in the Nelson-Siegel framework cannot be estimated accurately since the estimation relies on 17 observations only. The analysis based on the DNS–TVL–GARCH model provides estimates of  $\lambda$  using current and past observations. The resulting estimates are therefore based on more data and become more stable as a result. However, the DNS–TVL–GARCH specification is sufficiently flexible to provide an adequate representation of the changes in  $\lambda$  over time.

The dots in the graph of Panel (B) are the OLS estimates of the constant variance  $\sigma_t^2$  in the Nelson-Siegel model (4.2) with  $\lambda$  fixed at 0.0609 as in Diebold and Li (2006). The estimated common GARCH component of the DNS–TVL–GARCH model is also presented in this graph (with scale adjustment). It is encouraging that the estimated common GARCH component provides an accurate description of the time-varying volatility in the time series of yields. Deviations between the two estimates can be detected at the end of the 1980's.

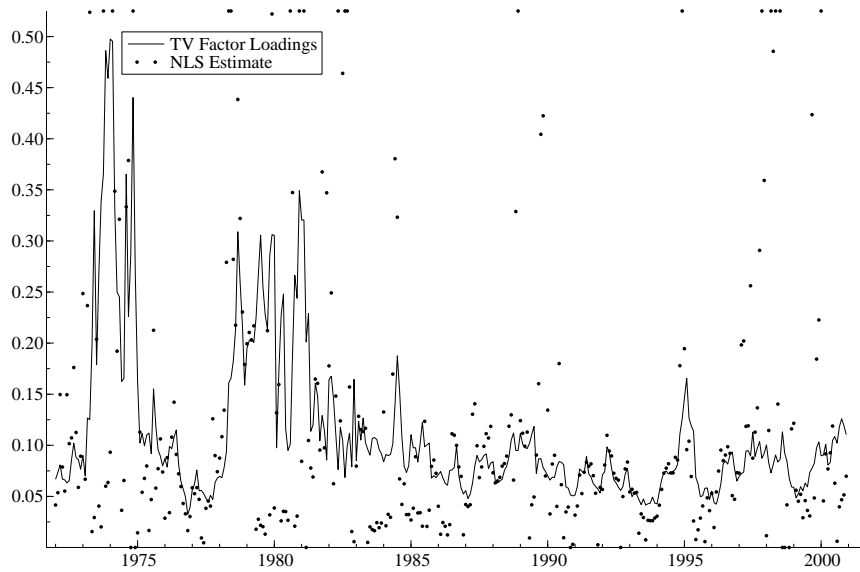
**(b) Results based on time-varying spline functions.** Our spline results as presented in Figure 4.2 provides some first indication that our empirical findings are not specific to a particular model specification. When more knots are chosen for the spline, the time-varying smooth functions become more flexible. The results presented in the figure are based on five knots which are equally spaced over the time-horizon of the sample. By increasing the number of knots, the time-varying  $\lambda$  estimates comes even closer to those obtained from the DNS-TVL model and displayed in Panel (A) of Figure 4.3. Similarly, by increasing the number of knots the spline volatility component comes closer to the estimated GARCH component as displayed in Panel (A) of Figure 4.4.

**(c) Results based on a different sample.** From the results presented in Panels

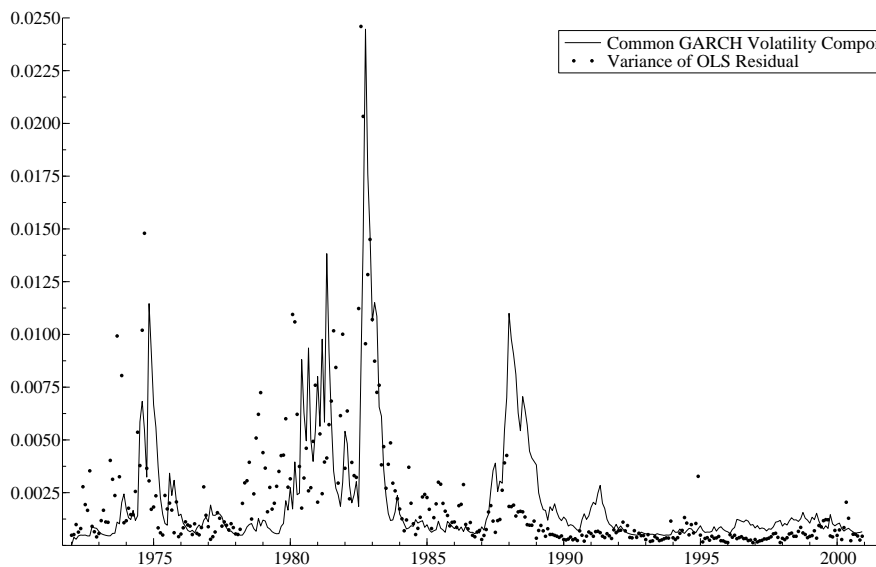
**Figure 4.7: Time-Varying Extensions compared to NLS and OLS Analysis**

In this figure we compare the time-varying factor loadings parameter and volatility component from the DNS-TVL-GARCH model to output from the NLS and OLS analysis. We compare the time-varying factor loadings parameter  $\lambda$  to estimates obtained from using NLS. The time-varying common GARCH volatility we compare to the residual variance from the OLS model.

(A) Time-Varying Factor Loadings, compared to NLS



(B) Common Time-Varying Volatility Component, compared to OLS





(A) of Figures 4.3 and 4.4 we have learned that the DNS–TVL–GARCH model particularly captures the variations in both  $\lambda$  and the volatility before 1987. It is therefore interesting to investigate whether the DNS–TVL–GARCH model also provides improvements in model fit for the data-set after 1987. For this purpose, we have re-estimated the baseline DNS model and its extensions for the sub-sample indicated by  $>1987$ . The results reported in Panel A of Table 4.4 are reproduced for the sub-sample  $>1987$  in the lower section of Panel B. We are encouraged by the empirical result that the model fit has increased for the TVL and GARCH extensions of the DNS model based on the  $>1987$  sample. The significant improvements for the GARCH extension of the DNS model are pronounced and most likely due to the volatility changes in the initial period after 1987 and in the middle of the 1990's.

## 4.5 An Illustration for an Unbalanced Dataset

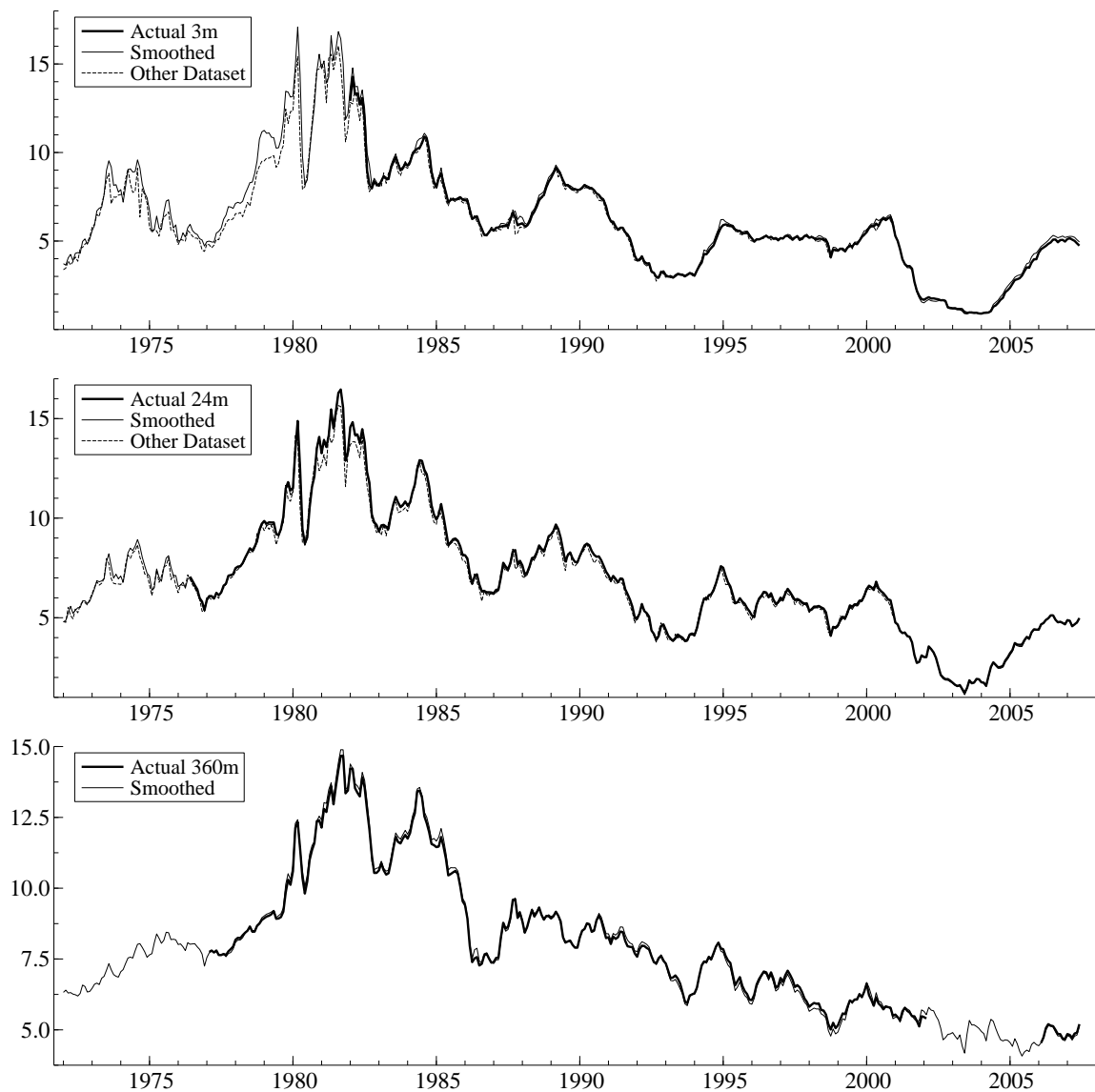
An attractive feature of models in state space form is that they can allow for missing values. For OLS estimation of the Nelson-Siegel model (as put forward by Diebold and Li (2006)) data must be available for all periods to avoid ad-hoc measures. However, this is not the case for models in state space form.

The smoothing algorithm associated with the Kalman filter produces the smoothed estimates of the latent factors for all periods and based on the available observations in the dataset. The estimation procedure itself does not change depending on data availability. Moreover, the smoothed estimates of the factors do also generate smoothed estimates of the interest rates for all maturities. It implies that when data is missing for a certain maturity, we are still able to obtain estimates for this maturity which are then based on data from this maturity at other time periods and data from other maturities. It is therefore expected that we still can obtain accurate forecasts for a maturity although the data is not complete for this maturity. This is a strong property of Kalman filter methods.

To illustrate this we use a publicly available dataset of fixed maturity U.S. Treasury yields. The dataset is obtained from the Federal Reserve Economic Data (FRED) online database, maintained by the Federal Reserve Bank of St. Louis. We look at the fixed maturity interest rates, over the period January 1972 up to June 2007, with maturities of 1, 3, 6, 12, 24, 36, 60, 84, 120, 240 and 360 months. With the convenience of being freely available and covering a long time horizon, there comes the inconvenience that many missing entries are present in the dataset. For example, the dataset for the 3 month bill starts only in January 1982, for the 24 month this is June 1976 and the 360 month starts February 1977 with missings in the period March 2002 until January 2006 (the period when it was not issued).

**Figure 4.8: An Illustration with Missing Values**

This figure illustrates how the Nelson-Siegel latent factor model deals with missing values. The data is the FRED fixed maturity U.S. Treasury Yields dataset from January 1972 up to June 2007 (note that all other tables and figures in this chapter are based on the unsmoothed Fama-Bliss data). We show the yield from the data, the smoothed yield using the Nelson-Siegel latent factor model and, if available, the yield from the unsmoothed Fama-Bliss dataset.



We have estimated the Nelson-Siegel latent factor model for this dataset. Figure 4.8 shows the time series of the maturities with missings that are mentioned above. We see that using the smoothed latent factors, based on data available not only in that period but also during other periods, an estimate of the missing yield can be obtained.

To obtain an indication of the reliability of this estimate, we compare it to its value obtained from the dataset of Section 4.4. Unfortunately, this dataset does not contain the 30 year bond and therefore we can only make such comparisons for the 3 month and 24 month yields. We find that in both cases the smoothed yield obtained provides a reliable estimate of the missing data.

## 4.6 Conclusion

The Nelson-Siegel framework provides means for an effective time series analysis of yield data. In this chapter we propose two extensions for the dynamic Nelson-Siegel (DNS) model of Diebold, Rudebusch, and Aruoba (2006) where the level, slope and curvature of the yield are treated as dynamic latent factors and modeled by a VAR process. The factor loadings in the DNS model depend on a single parameter that is usually taken as fixed. We show that the factor loading parameter can be estimated accurately from the data. It implies that the data can be highly informative about the factor loadings. Our first contribution concentrates on the question whether the factor loading parameter is constant over time. For this purpose we treat the loading parameter as the fourth latent factor in the DNS model. This nonlinear extension of the DNS model leads to a significant improvement in model fit. Next we turn our attention to the volatility pattern in each of the maturities and we focus on the question whether it is constant over time. For this purpose we introduce a common GARCH volatility component in the DNS model. The common volatility component is multiplied by a loading parameter for each maturity. The GARCH extension of the DNS model provides an even more significant improvement in model fit. The empirical results are obtained for a standard dataset that is analyzed by others in the literature. We have given evidence that our empirical results are robust against alternative model specifications and a different sample choice. Finally, we illustrate that missing values can be easily treated in this modeling framework. For example, we consider a dataset where four years of the 30 year bond has not been issued. However, with the use of both data from other periods of the same maturity and data from other maturities we obtain accurate estimates of the missing values. The general framework of the DNS model allows other modifications for future research.

# Chapter 5

## Conclusion

In this dissertation I look at various aspects of the market for U.S. treasuries. In the first part, Chapters 2 and 3, I look at trading of a futures contract that has as underlying U.S. government bonds with a long maturity. In the second part, Chapter 4, I study a model that is commonly used to analyze a large cross-section of bonds with different maturities over time. Each of the chapters in this dissertation contains a conclusion that summarizes the findings in each separate chapter. In the current chapter I highlight the most relevant results of each of the chapters and discuss how they relate to each other. I conclude with providing policy implications and directions for future research.

### 5.1 Trading U.S. Government Bonds

First I turn to a high-frequency dataset of the 30Y U.S. treasury futures that trades on the Chicago Board of Trade (CBOT). The data contains all trades for all floor traders on the trading pit over the period 1994-1997. I benefit from the unique feature of the data that the floor traders have to indicate whether a trade is for their own account or on behalf of a customer. This allows to filter out the part of order flow coming from customers. In addition this makes it possible to accurately track each trader's inventory position over time.

In Chapter 2 I find increased informativeness of customer order flow following macroeconomic announcements. I show that the floor traders need to observe customer flow to fully appreciate the just released news. This result adds support to recent papers that suggest order flow is informative beyond the classic equity market interpretation of containing a signal on a stock's future dividends. As I directly look at an end-user dataset the result is consistent with order flow containing private information at the micro level, about agents' preferences and endowments. In addition I show that traders that observe this customer flow are able to benefit from this information: they have higher profits from

trading.

In Chapter 3 I provide evidence against the classic notion that market makers are uninformed liquidity suppliers. Market makers initiate a large percentage of their trades, over 40%. In addition market makers initiate a significant percentage of their trades that increase their inventory position. The amount to which they do so relates positively and significantly to their profits from trading. This is the first direct evidence of active speculative position taking by market makers. The market making behavior is consistent with the Madhavan and Smidt (1993) model, where the market maker is both a speculator and a dealer.

I come across the challenge of signing trades in the absence of quotes. For markets with quotes there are algorithms available, such as the Lee and Ready (1991) algorithm. Hasbrouck (2004) suggests a Bayesian methodology for markets without quotes based on the Roll (1984) model. I propose an alternative to this method that analyzes the model with time series techniques for models in state space. Using a simulation study I show that this method has similar performance, but is over 14 times quicker.

## 5.2 Yields of U.S. Government Bonds

In Chapter 4 I focus on a long dataset with a cross-section of yields for different maturities. Recent advances in the econometric literature show the usefulness of the Nelson and Siegel (1987, NS) model to forecast the yield curve. Diebold and Li (2006) show that forecasts obtained using this model outperform competing economical and even statistical methods. Diebold, Rudebusch, and Aruoba (2006) argue that the NS model is in its nature a model in state space form, as it decomposes an observed time series into latent factors.

I suggest two extensions to the current approach with which researchers analyze the NS model. First, I relax the common assumption that the factor loadings parameter should be fixed over time. Second, I add time-varying volatility to the model, using a GARCH specification. The NS model with these extensions significantly improves the fit of the model. The model with time-varying volatility offers the great advantage that forecast confidence intervals are not constant over time. The model with time-varying factor loadings shows that the parameter governing these loadings is not constant. This provides evidence against both the assumption underlying the global yield curve model of Diebold, Li, and Yue (2007) and against an important implication of the Christensen, Diebold, and Rudebusch (2007) arbitrage-free Nelson-Siegel model.

### 5.3 Policy Implications

The results in this dissertation have several implications for policy. First, the high trading profits of floor traders that bring customer trades to the market (dual traders) compared to floor traders that do not do so (local traders) seems to offer evidence that customers get disadvantaged. However, the only variable with which I am able to measure this is profits from trading. As Fishman and Longstaff (1992) point out, in equilibrium no customer should be worse off than another provided customers are free to choose through which intermediary they trade. Fishman and Longstaff (1992) therefore argue that the customers of dual traders get other advantages, such as paying lower commissions or getting better execution. The results in this dissertation add to the dual trading literature by raising the argument that whether the customers of dual traders get disadvantaged depends on persistence in the role of floor traders. If there is no persistence, and duals switch roles to locals on certain days, customers do not really have a choice. With persistence it is indeed a question why customers trade through traders that act as duals. I find evidence against the common explanation that dual traders have superior trading skill, supporting the hypothesis that customers of dual traders receive other benefits.

Second, the results in this dissertation provide relevant insights into the role of market makers on financial exchanges. The recent trend in financial markets is toward electronic trading, but also here there are market participants that play the role of the ‘classic’ intermediaries and market makers that I examine. I show that market makers do not always supply liquidity, and can even be informed. In the design of a financial market it is important to realize that allowing market makers to speculate can reduce trading costs for other market participants. This as part of the fixed costs for making the market (such as the costs of having a ‘seat’ on the exchange) can then be paid for by the profit from speculation, whereas otherwise it all needs to come from customers.

Finally, my analysis of the dynamic Nelson and Siegel (1987) yield curve model is relevant for how central banks analyze the term structure. The extensions I propose improve the fit of the model. That I find the factor loadings parameter to be varying over time provides evidence against the assumption upon which a global yield curve is created, such as in Diebold, Li, and Yue (2007), and contradicts results from arbitrage-free Nelson-Siegel models, see Christensen, Diebold, and Rudebusch (2007).

### 5.4 Directions for Further Research

The results in this dissertation offer many directions for further research. Most notably, it is interesting to combine the two parts of the dissertation in various ways. In the first part of the dissertation I look at a high-frequency dataset of one maturity. In the second

part I look at a monthly dataset for a large group of maturities.

One direction for further research is to look at a large cross-section of maturities at a frequency higher than monthly. Doing so allows to look at many interesting subjects. First, one can look at the performance of the extended Nelson and Siegel (1987) model that I propose at higher frequencies than the monthly level. It is particularly interesting to see how the GARCH volatility behaves, which relates to the literature of aggregation of GARCH processes (see, e.g., Drost and Nijman (1993)). Second, in the first part of the dissertation I find macro news to be a significant determinant of trading variables such as volume and volatility. Analyzing the effects of macro news for a slightly lower frequency dataset (such as the daily level) for a cross-section of maturities provides further insights into the aggregation of macro information. Fleming and Remolona (1999b) derive a structural model for announcement impact curves. The flexibility of the Nelson and Siegel (1987) functional form provides an alternative way to model this. Combining a study of macroeconomic news and the term structure is also consistent with the direction for future research provided by Andersen and Benzoni (2008), who find evidence against the class of affine term structure models.

A separate direction for future research is to further develop the state space form regime switching methodology to sign trades. For now the model is relatively simple as to stay close to the celebrated Roll (1984) model. It is possible to follow Hasbrouck (2004) and extend the model in various directions. In addition the model can be used in other situations where there is a time series of prices but not of quotes.

# Samenvatting (Summary in Dutch)

## Risicovrije Rente Dynamiek: Informatie, Handel en State Space Modellen

De obligatiemarkt is de grootste financiële markt. Het totaal uitstaande volume is meer dan \$77 biljoen, vergeleken met \$60 biljoen voor de aandelenmarkt. Hiervan is meer dan 38% uit de Verenigde Staten afkomstig, waarvan weer \$6.5 biljoen is uitgegeven door de Amerikaanse overheid. In dit proefschrift richt ik me op verschillende aspecten van de markt voor staatsobligaties die zijn uitgegeven door de VS.

In het eerste deel van dit proefschrift, hoofdstukken 2 en 3, richt ik me op de handel in deze staatsobligaties. Academische studies die kijken naar de trade-by-trade handel in financiële waarden en de kosten die hieraan verbonden zijn vallen in het deelgebied van de financiële economie die vaak wordt aangeduid als market microstructure. Deze term komt uit Garman (1976), die het moment tot moment marktgedrag behandelt. Sinds dit artikel is de literatuur uit dit deelgebied snel gegroeid, vooral sinds snelle computers beschikbaar zijn waarmee grote datasets geanalyseerd kunnen worden.

Een van de belangrijkste onderwerpen in de market microstructure is het verschil tussen de prijs waartegen gekocht (de laatkoers) en verkocht (de biedkoers) kan worden. Het verschil tussen deze twee, de zogenaamde spread, is een indicatie van de kosten van het handelen. Partijen die bij de markt komen om te kopen of verkopen zijn partijen die liquiditeit vereisen, en betalen deze kosten. Handelaren die klaar staan om deze transacties te accommoderen leveren liquiditeit. Drie factoren bepalen hoe groot de spread is die betaald wordt: kosten voor het verwerken van de order (Roll (1984)), kosten voor het dragen van een voorraad van de financiële waarde (Stoll (1978)) en kosten voor informatie die asymmetrisch verdeeld is onder de markt participanten (Kyle (1985) en Glosten en Milgrom (1985)). Orderverwerkingskosten zijn de kosten die liquiditeit verschaffers rekenen behalve de andere twee genoemde kosten. Hier zitten onder andere de kosten van arbeid en kapitaal bij. De liquiditeit verschaffer rekent voorraadkosten omdat hij door het aangaan van een transactie afwijkt van zijn optimale positie. Asymmetrische informatie kosten komen voor uit de kans dat de tegenpartij in een transactie wellicht meer infor-



matie heeft over de financiële waarde en dat de liquiditeit verschaffer geld verliest omdat hij aan de verkeerde kant van de transactie zit.

Het kan erop lijken dat deze handelskosten beperkt zijn, wellicht niet genoeg om een hele stroom wetenschappelijke studies te verantwoorden of zelfs dit proefschrift. Het is inderdaad waar dat de spread beperkt is. Kijkende naar de Amerikaanse markt voor staatsobligaties is dit slechts \$6 per transactie. Echter, gegeven het enorme volume op de financiële markten tellen deze kleine bedragen snel op. Dit wordt duidelijk als we de markt voor Amerikaanse staatsobligaties bekijken: met meer dan 20.000 transacties op een dag van gemiddeld 13 contracten per transactie is het bedrag al \$1,56 miljoen per dag. Dit is \$7.8 miljoen per week en meer dan \$390 miljoen per jaar. En deze bedragen zijn alleen voor de markt voor Amerikaanse staatsobligaties, het totaalbedrag van deze kosten voor alle financiële markten gecombineerd geeft een getal in de miljard dollar!

In hoofdstuk 2 kijk ik naar de derde uitleg voor de spread, informatie die asymmetrisch verdeeld is onder de markt participanten. Ik richt mij in het bijzonder op de handel in Amerikaanse staatsobligaties nadat macro-economisch nieuws wordt vrijgegeven. In de Verenigde Staten wordt dit nieuws op vooraf aangekondigde tijdstippen bekend gemaakt. Het is bijvoorbeeld bekend dat iedere eerste vrijdag van de maand om exact 8:30 ET het Amerikaanse Bureau of Labor Statistics het werkloosheidsrapport vrijgeeft. Deze nieuwsberichten veranderen de risicovrije rentevoet. Veel artikelen rapporteren dat de macro-economische nieuwsberichten verantwoordelijk zijn voor de meeste volatiliteit patronen over de dag (zie bijvoorbeeld Ederington en Lee (1993, 1995), Balduzzi, Elton en Green (2001), Andersen, Bollerslev, Diebold en Vega (2003, 2007) en Fleming en Remolona (1997, 1999a)). Green (2004) vindt dat er meer informatie zit in de stroom van transacties (de zogenaamde order flow) na dit macro nieuws. Door te kijken naar transacties die hebben plaatsgevonden kunnen participanten meer geïnformeerd raken over wat de werkelijke evenwicht risicovrije rentevoet hoort te zijn. Deze resultaten tonen aan dat informatie meer asymmetrisch is verdeeld over de economie nadat er macro nieuws is vrijgekomen, en dat order flow de evenwicht risicovrije rentevoet significant beïnvloedt.

Het is niet eenvoudig om dit resultaat te interpreteren. In de klassieke literatuur voor aandelenmarkten (zoals Glosten en Milgrom (2004) en Kyle (1985)) wordt informatie in order flow vaak geïnterpreteerd als een teken dat sommige markt participanten meer informatie hebben over toekomstige dividenden van een bedrijf. Er is echter recent theoretisch en empirisch werk dat aangeeft dat informatie in order flow breder kan zijn dan enkel deze uitleg. Hasbrouck en Seppi (2001) vinden een gemeenschappelijke component in order flow van verschillende bedrijven. Deze gezamenlijke component beïnvloedt de markt returns. Edelen en Warner (2001) leveren meer bewijs voor dit, en tonen aan dat geldstromen in en uit beleggingsfondsen een dergelijke gezamenlijke component is. Ook

in financiële markten anders dan de aandelenmarkt is er bewijs dat er meer informatie in order flow zit dan slechts aandelenspecifieke informatie. Behalve het bovengenoemde werk van Green (2004) voor de markt voor staatsobligaties (dat verder is gedocumenteerd door o.a. Brandt en Kavajecz (2004) en Pasquariello en Vega (2006)) is er bewijs voor de wisselkoersmarkt (zie bijvoorbeeld Evans en Lyons (2002, 2008)). Recente theorieën geven aan dat order flow informatie kan bevatten op het micro niveau van de economie, over bezittingen en risico preferenties van markt participanten (zie Gallmeyer, Hollifield en Seppi (2005) en Saar (2007)).

Met een unieke dataset ben ik in staat om het deel van de totale order flow te herleiden dat voortkomt uit de aggregatie van micro informatie. Ik bestudeer de futures markt voor Amerikaanse staatsobligaties met een lange looptijd van 30 jaar. Het volume van de handel voor deze looptijd concentreert zich voor 95% op de futures markt (in plaats van de spot markt). We bekijken deze markt in de periode 1994-1997. In deze periode was elektronische handel in dit instrument nog zeer beperkt, en vond praktisch alle handel plaats op de beursvloer. Als klanten van buiten de vloer wilden handelen moest dit door een van de handelaren op de vloer (zogenaamde floor traders) te benaderen. Vanwege regelgeving moesten de handelaren op de beursvloer aangeven of een transactie voor zichzelf was of voor klanten. De dataset die ik in dit proefschrift gebruik bevat alle transacties van alle floor traders. Doordat ik mij in hoofdstuk 2 alleen richt op het klantendeel van de order flow bestudeer ik direct een dataset die micro informatie betreft. Dit staat in contrast met de andere genoemde studies die kijken naar de markt voor staatsobligaties. Deze artikelen bestuderen datasets die zich richten op transacties die plaatsvinden tussen handelaren. Deze zogenaamde interdealer order flow kan gerelateerd zijn aan transacties van klanten, maar kan verschillen door twee redenen. Allereerst is het mogelijk dat handelaren hun transacties maken naar aanleiding van informatie in de order flow van klanten. Dit kan de waargenomen informativiteit van order flow beïnvloeden. Ten tweede kunnen handelaren hun transacties maken naar aanleiding van informatie die ze zelf hebben, los van hun klantenorders (dit werd bijvoorbeeld gevonden door Anand en Subrahmanyam (2007)).

In hoofdstuk 2 laat ik zien dat order flow komende van klanten inderdaad informatief is voor het herleiden van de nieuwe evenwicht risicovrije rentevoet. Order flow komende van klanten is significant meer informatief in het 15 minuten interval na macro economisch nieuws. Deze significantie is zowel statistisch als economisch van aard. Ik vind dat 25% van de verklaarde variatie in de risicovrije voet na macro nieuws komt van klanten order flow. Hiernaast laat ik zien dat de handelaren op de beursvloer die een deel van de klanten order flow kunnen zien deze informatie gebruiken om te speculeren. Ik vind dat handelwinsten voor dual traders, floor traders die zowel voor zichzelf als voor klanten handelen,

hoger zijn dan die van local traders, floor traders die alleen voor zichzelf handelen. Dit resultaat houdt ook stand in de cross-sectie van dual traders: handelswinsten zijn positief en significant gerelateerd aan diverse maatstaven voor toegang tot klantenorders.

In hoofdstuk 3 kijk ik naar de tweede uitleg voor de spread, voorraadkosten voor de liquiditeit verschaffer die voortkomen uit het risico van een suboptimale positie in een financiële waarde. De grote cross-sectie van handelaren die er actief is op de futures markt voor Amerikaanse staatsobligaties met tot meer of mindere mate een rol van liquiditeit verschaffer geeft een ideale sample om dit nader te onderzoeken. Handelaren die zich bezig houden met het verschaffen van liquiditeit worden vaak aangeduid als market makers (de hoekman op de Nederlandse beurs was hier een voorbeeld van). Deze personen handelen voor zichzelf: ze kopen als er verkopers op de beurs arriveren en hopen deze aandelen later dan weer te verkopen aan kopers (en omgekeerd). Het risico dat ze lopen doordat ze een tijdje met (of zonder) het aandeel zitten wordt vergoed door de spread. Tot nu toe is er echter beperkt empirisch bewijs voor modellen die het gedrag van market makers modelleren zoals hierboven omschreven, ondanks de vele artikelen die zich hierop hebben gericht. Een mogelijke verklaring hiervoor is dat deze liquiditeit verschaffers soms zelf speculeren. Madhavan en Smidt (1993) ontwikkelen een model waar transacties van de market makers uit twee componenten bestaan. Allereerst handelen ze zoals verwacht: ze initiëren transacties die hun voorraad terugbrengen naar hun een lange termijn optimale positie. De tweede component relateert echter tot korte termijn speculatieve belangen.

Ik gebruik dezelfde dataset als in hoofdstuk 2 en toon aan dat de market makers op de futures markt ook speculatieve posities opbouwen. Ik vind bewijs voor beide componenten van het Madhavan en Smidt (1993) model. De voorraadpositie keert snel terug naar een lange termijn optimale positie: de floor traders sluiten de dag af met een positie rond 0 obligaties. Daarnaast vind ik dat floor traders transacties initiëren die hun voorraadpositie wegbrengt van dit lange termijn optimum. Als ik de mate waarin dit gebeurt relateer aan handelswinst vind ik een positieve en significante relatie. Deze resultaten dragen bij aan het debat of de market maker inderdaad ‘simpelweg’ een liquiditeit verschaffer is zonder privé informatie of een geïnformeerde speculant. Artikelen zijn tot nu toe tot verschillende conclusies gekomen, allen leveren echter indirect bewijs. Het derde hoofdstuk in dit proefschrift levert direct bewijs van winstgevende speculatie van market makers.

In de analyse van hoofdstuk 3 loop ik tegen de uitdaging aan van het signen van transacties. Hierin wordt voor iedere transactie bekeken of de kopende of de verkopende partij de transactie heeft geïnitieerd. Voor markten waar er quote informatie beschikbaar is (de bied- en laatkoers) zijn er algoritmes beschikbaar (zie bijvoorbeeld Lee en Ready (1991)). Deze algoritmes relateren de transactieprijs aan het gemiddelde van de bied- en laatkoers (de zogenaamde midquote). Het signen van de transacties wordt echter (nog)

meer uitdagend wanneer quote informatie niet beschikbaar is omdat er dan geen midquote beschikbaar is om de transactieprijs mee te vergelijken. Hasbrouck (2004) ontwikkelt een Bayesiaans algoritme voor deze taak, gebaseerd op het Roll (1984) model. Echter, omdat dit op simulaties is gebaseerd is dit numeriek gezien een erg duur algoritme. Met meer dan 40 miljoen transacties in mijn dataset duurt dit algoritme te lang. Ik ontwikkel een alternatieve implementatie van het Roll (1984) model, gebaseerd op technieken uit de tijdreeksanalyse voor modellen in state space form. In dit soort modellen wordt een (mogelijk multivariate) tijdreeks verklaard door latente (niet waargenomen) factoren. Onder de aanname van normaal verdeelde ruis en een normaal verdeeld transitie proces voor deze niet waargenomen factoren kan het Kalman filter gebruikt worden om schattingen te krijgen van de parameters en de latente factoren. Het Roll (1984) model past goed in het framework van deze modellen: de waargenomen prijs wordt verklaard met een niet waargenomen ‘echte’ prijs en een tijdreeks van discrete transactie signs. Het model is echter niet-lineair, om toch schattingen te verkrijgen volg ik de macro-economische literatuur over regime switching modellen. Met simulatie toon ik aan dat het resulterende algoritme meer dan 10 keer sneller is dan Hasbrouck’s (2004) algoritme en even accuraat is.

In het tweede deel van dit proefschrift, hoofdstuk 4, richt ik me op de termijn structuur van de risicovrije rentevoet: de cross-sectie van rentes met een verschillende looptijd. De unieke structuur van de termijn curve houdt onderzoekers al lang bezig. Rentes met een verschillende looptijd zijn gerelateerd aan elkaar: rentes op obligaties met een lange looptijd kunnen worden gezien als risicogecorrigeerde verwachtingen van toekomstige rentes op obligaties met een kortere looptijd. Het is hierdoor mogelijk om de informatie in de termijn structuur samen te vatten met een aantal gemeenschappelijke factoren die de dynamiek van de rente samenvatten. Er is echter nog veel onbekend over de exacte relatie tussen de rentes met een verschillende looptijd en wat precies de markt voor staatsobligaties drijft. Het lijkt erop dat het model dat nog steeds het beste werkt om de termijn structuur te voorspellen de random walk is: de voorspelling van de termijn structuur van volgende maand is de huidige termijn structuur.

Er is veel belang bij het beter begrijpen van de termijn structuur. Piazzesi (2003) noemt vier redenen waarom onderzoekers naar de termijn structuur moeten kijken. Allereerst bevat de termijn structuur informatie over toekomstige economische ontwikkelingen door de relatie tussen staatsobligaties met een korte en lange looptijd. Ten tweede is er belang voor monetair beleid. In veel landen kan de centrale bank alleen de korte termijn rente beïnvloeden, terwijl voor veel beslissingen van consumenten de lange termijn rente het meest belangrijk is. Het begrijpen van de relatie tussen de korte en lange termijn rente zal beleidsbepalers helpen hun beslissingen te nemen. Ten derde is het voor de overheid

van belang om te weten hoe de hele termijn structuur reageert als er staatsobligaties van een bepaalde looptijd worden uitgegeven. Tot slot is de termijn structuur van belang voor het prijzen van derivaten en voor hedging. De prijs van veel derivaten hangt af van de gehele termijn structuur. Banken proberen het risico dat ze lopen vanwege hun korte termijn rente betalingen en lange termijn rente ontvangsten af te dekken.

De eerste onderzoeken naar de termijn structuur zijn allen theoretisch van aard. In de vroege modellen van Vasicek (1977) en Cox, Ingersoll en Ross (1985) verklaart een niet waargenomen rente op een oneindig korte termijn de hele termijn structuur. Deze klasse van modellen werd later uitgebreid tot meer factoren. In de jaren '90 kwam een tweede belangrijke ontwikkeling, toen Hull en White (1990) en Heath, Jarrow en Morton (1990) een klasse van rente modellen ontwikkelden die arbitrage mogelijkheden op basis van de termijn structuur uitsluit. Zeer recent is er een vernieuwde interesse in het Nelson en Siegel (1987, hierna NS) model voor de termijn structuur. Deze methode verklaart de hele termijn structuur met drie niet waargenomen componenten die worden geïnterpreteerd als de level (niveau), slope (helling) en curvature (krommigheid) van de termijn structuur. De methode is populair bij zowel mensen in de financiële wereld als centrale bankiers (9 van de 13 ondervraagde centrale banken in een studie van de Bank of International Settlements gebruiken dit model om de termijn structuur te schatten). Het grote voordeel is dat het model statistisch van aard is, maar toch een heldere interpretatie heeft. Twee recente publicaties van Francis X. Diebold hebben het belang van het NS model versterkt. Diebold en Li (2006) laten zien dat voorspellingen verkregen met het NS model die van concurrerende economische en statistische modellen verbeteren. Diebold, Rudebusch en Aruoba (2006) beargumenteren dat het NS model in essentie een model in state space form is, aangezien een waargenomen tijdreeks verklaard wordt met latente factoren.

In hoofdstuk 4 onderzoek ik verschillende aannames die gemaakt worden om het Nelson en Siegel (1987) model te kunnen schatten. Allereerst wordt er aangenomen dat de parameter die bepaalt hoe de factoren de termijn structuur beïnvloeden constant is. Ik stel een extensie van het model voor waarin deze parameter tijdsvariërend is. Allereerst doe ik dit door deze parameter te modelleren als een spline functie. Aangezien deze analyse aangeeft dat deze parameter inderdaad sterk varieert over tijd modelleer ik het op een nog flexibelere manier, door het toe te voegen aan het model als een latente factor. Op deze manier wordt deze parameter behandeld als de level, slope en curvature factoren. De niet-lineairiteit die hieruit volgt los ik op door het model te analyseren met behulp van de extended Kalman filter. De tweede aanname die ik bekijk is die van constante volatiliteit. Ook deze modelleer ik eerst met een spline functie, waarmee ik vind dat ook deze sterk varieert over de tijd. Gemotiveerd door dit resultaat bestudeer ik dit nader, met een GARCH functie zoals in Engle (1982) en Bollerslev (1986).

Ik toon aan dat het loslaten van deze twee aannames het model significant verbetert, zowel vanuit een statistisch als een economisch oogpunt. Bovendien geeft het tijdsvariërende model inzichten in enkele nieuwe ontwikkelingen in de literatuur. Christensen, Diebold en Rudebusch (2007) ontwikkelen een NS model dat arbitrage uitsluit. Een conditie voor hun model is echter constantheid van de parameter die bepaalt hoe de factoren de termijn structuur beïnvloeden. Diebold, Li en Yue (2007) ontwikkelen een termijn structuur model waarin rentes uit verschillende landen tegelijkertijd worden bekeken. Deze is echter ook gebaseerd op constantheid van de genoemde parameter. De resultaten in hoofdstuk 4 van dit proefschrift leveren bewijs hiertegen.



# Bibliography

- Anand, A., and A. Subrahmanyam. 2007. "Information and the Intermediary: Are Market Intermediaries Informed Traders in Electronic Markets?" *Journal of Financial and Quantitative Analysis* (forthcoming).
- Anand, A., C. Tanggaard, and D.G. Weaver. 2008. "Paying for Market Quality." Working Paper, WFA 2007 paper.
- Andersen, T.G., and L. Benzoni. 2008. "Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models." AFA 2007 Paper.
- Andersen, T.G., T. Bollerslev, F.X. Diebold, and C. Vega. 2003. "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange." *American Economic Review* 93:p38–62.
- . 2007. "Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets." *Journal of International Economics* 73:p251–277.
- Anderson, B.D.O., and J.B. Moore. 1979. *Optimal Filtering*. Englewood Cliffs:PrenticeHall.
- Ang, A., and M. Piazzesi. 2003. "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics* 50:p745–787.
- Ansley, C.F., and R. Kohn. 1986. "A Note on Reparameterizing a Vector Autoregressive Moving Average Model to Enforce Stationarity." *J. Statistical Computation and Simulation* 24:p99–106.
- Balduzzi, P., E.J. Elton, and T.C. Green. 2001. "Economic News and Bond Prices: Evidence From the U.S. Treasury Market." *Journal of Financial and Quantitative Analysis* 36:p523–543.
- Benveniste, L.M., A.J. Marcus, and W.J. Wilhelm. 1992. "What's Special about the Specialist?" *Journal of Financial Economics* 32:p61–86.



- Bessembinder, H., J. Hao, and M. Lemmon. 2007. "Why Designate Market Makers? Affirmative Obligations and Market Quality." Working Paper.
- Biais, B., L. Glosten, and C. Spatt. 2005. "A Survey of Microfoundations, Empirical Results, and Policy Implications." *Journal of Financial Markets* 8:p217–264.
- Bianchi, F., H. Mumtaz, and P. Surico. 2006. "The UK Great Stability: a View from the Term Structure of Interest Rates." Working Paper.
- Bjønnes, G.H., and D. Rime. 2005. "Dealer Behavior and Trading Systems in Foreign Exchange Markets." *Journal of Financial Economics* 75:p571–605.
- Bliss, R.R. 1997. "Testing Term Structure Estimation Methods." *Advances in Futures and Options Research* 9:p197–231.
- Bollerslev, T. 1986. "Generalised Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 51:p307–327.
- Boulatov, A., and T.J. George. 2007. "Securities Trading when Liquidity Providers are Informed." Working Paper, AFA 2008 paper.
- Bowsher, C.G., and R. Meeks. 2008. "High Dimensional Yield Curves: Models and Forecasting." *Journal of the American Statistical Association (forthcoming)*.
- Brandt, M.W., and K.A. Kavajecz. 2004. "Price Discovery in the U.S. Treasury Market: The Impact of Order Flow and liquidity on the Yield Curve." *Journal of Finance* 59:p2623–2654.
- Cao, H.H., M.D. Evans, and R.K. Lyons. 2006. "Inventory Information." *Journal of Business* 79:p325–363.
- Chakravary, S., and K. Li. 2003. "An Examination of Own Account Trading by Dual Traders in Futures Markets." *Journal of Financial Economics* 69:p375–397.
- Chan, K. 1992. "A Further Analysis of the Lead-Lag Relationship between the Cash Market and Stock Index Futures Market." *Review of Financial Studies* 5:p123–152.
- Chang, E.C., P.R. Locke, and S.C. Mann. 1994. "The Effect of CME Rule 552 on Dual Traders." *Journal of Futures Markets* 14:p493–510.
- Christensen, J.H.E., F.X. Diebold, and S.B. Rudebusch. 2007. "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models." Manuscript University of Pennsylvania.
- Chung, K.H., C. Chuwonganant, and D.T. McCormick. 2004. "Order Preferencing and Market Quality on NASDAQ Before and After Decimalization." *Journal of Financial Economics* 71:p581–612.

- Cox, R., J. Ingersoll, and S. Ross. 1985. "A Theory of the Term Structure of Interest Rates." *Econometrica* 53 (March): p385–407.
- Dai, Q., and K.J. Singleton. 2000. "Specification Analysis of affine term structure models." *Journal of Finance* 55 (5): p1943–1978.
- . 2002. "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of The Term Structure." *Journal of Financial Economics* 63:p415–441.
- Demsetz, H. 1968. "The Cost of Transacting." *Quarterly Journal of Economics* 82:p33–53.
- De Pooter, M. 2007. "Examining the Nelson-Siegel Class of Term Structure Models." Working Paper.
- De Pooter, M., F. Ravazzolo, and D. Van Dijk. 2007. "Predicting the Term Structure of Interest Rates." TI Discussion Paper.
- Diebold, F.X., and C. Li. 2006. "Forecasting the Term Structure of Government Bond Yields." *Journal of Econometrics* 130:p337–364.
- Diebold, F.X., C. Li, and V. Yue. 2007. "Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach." *Journal of Econometrics* (forthcoming).
- Diebold, F.X., S.B. Rudebusch, and S.B. Aruoba. 2006. "The Macroeconomy and the Yield Curve." *Journal of Econometrics* 131:p309–338.
- Doornik, J.A. 1998. *Object-oriented Matrix Programming Using Ox 2.0*. Timberlake Consultants Ltd.
- Driessen, J. 2005. "Is Default Risk Priced in Corporate Bonds?" *Review of Financial Studies* 18:p165–195.
- Drost, F.C., and T.E. Nijman. 1993. "Temporal Aggregation of Garch Processes." *Econometrica* 61:p909–927.
- Duffee, G.R. 2002. "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance* 57:p405–443.
- Duffie, D., and R. Kan. 1996. "A Yield-Factor Model of Interest Rates." *Journal of Mathematical Finance* 6 (2): p379–406.
- Duncan, D.B., and S.D. Horn. 1972. "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis." *Journal of the American Statistical Association* 67:p815–21.
- Dunne, P., H. Hau, and M. Moore. 2007. "A Tale of Two Platforms: Inter-Dealer and Retail Quotes in the European Bond Markets." Working Paper.

- Durbin, J., and S.J. Koopman. 2001. *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series.
- Easley, D., N.M. Kiefer, and M. O'Hara. 1996. "Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow." *Journal of Finance* 51:p811–833.
- Edelen, R.M., and J.B. Warner. 2001. "Aggregate Price Effects of Institutional Trading: A Study of Mutual Fund Flow and Market Returns." *Journal of Financial Economics* 59:p195–220.
- Ederington, L., and J. Lee. 1993. "How Markets Process Information: News Releases and Volatility." *Journal of Finance* 45:p1161–1191.
- . 1995. "The Short-Run Dynamics of the Price Adjustment to New Information." *Journal of Financial and Quantitative Analysis* 30:p117–134.
- Ellis, K., R. Michaely, and M. O'Hara. 2000. "The Accuracy of Trade Classification Rules: Evidence from Nasdaq." *Journal of Financial and Quantitative Analysis* 35:p529–552.
- Engle, R.F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50:p987–1007.
- Engle, R.F., V.K. Ng, and M. Rothschild. 1990. "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills." *Journal of Econometrics* 45:p213–238.
- Evans, M.D.D., and R.K. Lyons. 2002. "Order Flow and Exchange Rate Dynamics." *Journal of Political Economy* 110:p170–180.
- . 2005. "Meese-Rogoff Redux: Micro-Based Exchange Rate Forecasting." *American Economic Review* 95:p405–414.
- . 2007. "Exchange Rate Fundamentals and Order Flow." Working Paper.
- . 2008. "How is Macro News Transmitted to Exchange Rates?" *Journal of Financial Economics* 88:p26–50.
- Fama, E.F., and R.R. Bliss. 1987. "The Information in Long-Maturity Forward Rates." *American Economic Review* 77:p680–692.
- Fishman, J.F., and F.A. Longstaff. 1992. "Dual Trading in Futures Markets." *Journal of Finance* 47:p643–671.
- Fleming, M.J. 1997. "The Round-the-Clock Market for U.S. Treasury Securities." *Economic Policy Review* 97-July:p9–32.
- Fleming, M.J., and M. Piazzesi. 2005. "Monetary Policy Tick-by-tick." Working Paper.

- Fleming, M.J., and E.M. Remolona. 1997. "What Moves the Bond Market?" *Economic Policy Review* 97-Dec:p31–50.
- . 1999a. "Price Formation and Liquidity in the U.S. Treasury Market: The Response to Public Information." *Journal of Finance* 54:p1901–1915.
- . 1999b. "The Term Structure of Announcement Effects." Federal Reserve Bank of New York Staff Reports, no. 76.
- Fleming, M.J., and A. Sarkar. 1999. "Liquidity in U.S. Treasury Spot and Futures Markets." *Market Liquidity Research Findings and Selected Policy Implications*.
- Frino, A., and E. Jarnecic. 2000. "An Empirical Analysis of the Supply of Liquidity by Locals in Futures Markets." *Pacific-Basin Finance Journal* 8:p443–456.
- Gallmeyer, M., B. Hollifield, and D.J. Seppi. 2005. "Demand Discovery and Asset Pricing." Working Paper.
- Garbade, K.D., and J.F. Ingber. 2005. "The Treasury Auction Process: Objectives, Structure, and Recent Adaptations." *Current Issues in Economics and Finance* 11:p1–11.
- Garleanu, N., L.H. Pedersen, and A.M. Poteshman. 2005. "Demand-Based Option Pricing." NBER Working Paper, EFA 2005 Paper.
- Garman, M. 1976. "Market Microstructure." *Journal of Financial Economics* 3:p257–275.
- Glosten, L.R., and L.E. Harris. 1988. "Estimating the Components of the Bid-Ask Spread." *Journal of Financial Economics* 21:p123–142.
- Glosten, L.R., and P.R. Milgrom. 1985. "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *Journal of Financial Economics* 14:p71–100.
- Green, R.C., B. Hollifield, and N. Schürhoff. 2007. "Financial Intermediation and the Costs of Trading in an Opaque Market." *Review of Financial Studies* 20:p275–315.
- Green, T.C. 2004. "Economic News and the Impact of Trading on Bond Prices." *Journal of Finance* 59:p1201–1233.
- Grossman, S.J. 1989. "An Economic Analysis of Dual Trading." Rodney L. White Center for Financial Research Paper.
- Hamilton, J. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57:357–84.

- Hansch, O., N.Y. Naik, and S. Viswanathan. 1998. "Do Inventories Matter in Dealership Markets? Evidence from the London Stock Exchange." *Journal of Finance* 53:p1623–1656.
- Harris, L.E. 1990. "Statistical Properties of the Roll Serial Covariance Bid-Ask Spread Estimator." *Journal of Finance* 45:p579–590.
- . 2003. *Trading & Exchanges*. Oxford University Press.
- Harvey, A.C. 1989. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
- Harvey, A.C., E. Ruiz, and E. Sentana. 1992. "Unobserved Component Time Series Models with ARCH Disturbances." *Journal of Econometrics* 52:p129–158.
- Hasbrouck, J. 1988. "Trades, Quotes, Inventories, and Information." *Journal of Financial Economics* 22:p229–252.
- . 1999. "The Dynamics of Discrete Bid and Ask Quotes." *Journal of Finance* 54:p2109–2142.
- . 2004. "Liquidity in the Futures Pit: Inferring Market Dynamics from Incomplete Data." *Journal of Financial and Quantitative Analysis* 39:p305–326.
- . 2007. *Empirical Market Microstructure*. Oxford University Press.
- Hasbrouck, J., and D.J. Seppi. 2001. "Common Factors in Prices, Order Flows, and Liquidity." *Journal of Financial Economics* 59:383–411.
- Hasbrouck, J., and G. Sofianos. 1993. "The Trades of Market Makers: An Empirical Examination of NYSE Specialists." *Journal of Finance* 48:p1565–1593.
- Heath, D., R. Jarrow, and A. Morton. 1990. "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation." *Journal of Financial and Quantitative Analysis* 25:p419–440.
- . 1992. "Bond Pricing and the Term Structure of Interest Rates: a New Methodology for Contingent Claims Valuation." *Econometrica* 60:p77–105.
- Hendershott, T., C.M. Jones, and A.J. Menkveld. 2007. "Does Algorithmic Trading Improve Liquidity?" Working Paper.
- Hendershott, T., and M. Seasholes. 2007. "Market Maker Inventories and Stock Prices." *American Economic Review* 97:p210–214.
- Hull, J., and A. White. 1990. "Pricing Interest-Rate-Derivative Securities." *Review of Financial Studies* 3:p573–592.
- Jain, P. 2005. "Institutional Design and Liquidity on Stock Exchanges around the World." Working Paper.

- Kawaller, I.G., P.D. Koch, and T.W. Koch. 1987. "The Temporal Price Relationship Between S&P 500 Futures and the S&P500 Index." *Journal of Finance* 42:p1309–1329.
- Kim, C. J. 1994. "Dynamic Linear Models with Markov-Switching." *Journal of Econometrics* 60:p1–22.
- Kim, C. J., and C. R. Nelson. 1999. *State Space Models with Regime Switching*. Cambridge, Massachusetts: MIT Press.
- Kim, O., and R.E. Verrecchia. 1994. "Market Liquidity and Volume around Earnings Announcements." *Journal of Accounting and Economics* 17:p41–67.
- . 1997. "Pre-Announcement and Event-Period Private Informations." *Journal of Accounting and Economics* 24:p395–419.
- Koopman, S.J., M. Mallee, and M. Van der Wel. 2007. "Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters." *Journal of Business and Economic Statistics* (forthcoming).
- Koopman, S.J., N. Shephard, and J.A. Doornik. 1999. "Statistical Algorithms for Models in State Space using SsfPack 2.2." *Econometrics Journal* 2:p113–166.
- Kurov, A., and D.J. Lasser. 2004. "Price Dynamics in the Regular and E-Mini Futures Markets." *Journal of Financial and Quantitative Analysis* 39:p365–384.
- Kyle, A.S. 1985. "Continuous Auctions and Insider Trading." *Econometrica* 53:p1315–1336.
- Lee, C.M., and M.J. Ready. 1991. "Inferring the Trade Direction from Intraday Data." *Journal of Finance* 46:p733–746.
- Locke, P.R., and P. Sarajoti. 2004. "Aggressive Dealer Pricing." *The Quarterly Review of Economics and Finance* 44:p559–573.
- Lyons, R. 1995. "Tests of Microstructural Hypotheses in the Foreign Exchange Market." *Journal of Financial Economics* 39:p321–351.
- . 1997. "A Simultaneous Trade Model of the Foreign Exchange Hot Potato." *Journal of International Economics* 42:p275–298.
- MacKinlay, A.C., and K. Ramaswamy. 1988. "Index-Futures Arbitrage and the Behavior of Stock Index Futures Prices." *Review of Financial Studies* 1:p137–158.
- Madhavan, A. 2000. "Market Microstructure: A Survey." *Journal of Financial Markets* 3:p205–258.
- Madhavan, A., and S. Smidt. 1991. "A Bayesian Model of Intraday Specialist Pricing." *Journal of Financial Economics* 30:p99–134.

- . 1993. “An Analysis of Changes in Specialist Inventories and Quotations.” *Journal of Finance* 48:p1595–1628.
- Madhavan, A., and G. Sofianos. 1998. “An Empirical Analysis of NYSE Specialist Trading.” *Journal of Financial Economics* 48:p189–210.
- Madrigal, V. 1996. “Non-Fundamental Speculation.” *Journal of Finance* 51:p553–576.
- Manaster, S., and S.C. Mann. 1996. “Life in the Pits: Competitive Market Making and Inventory Control.” *Review of Financial Studies* 9:p953–975.
- . 1998. “Sources of Market Making Profits: Man Does Not Live by Spread Alone.” Working Paper.
- Menkveld, A.J., S.J. Koopman, and A. Lucas. 2007. “Modelling Round-the-Clock Price Discovery for Cross-Listed Stocks using State Space Methods.” *Journal of Business and Economic Statistics* 25:p213–225.
- Menkveld, A.J., A. Sarkar, and M. Van der Wel. 2007. “Discovering the Equilibrium Riskfree Rate.” Tinbergen Institute Discussion Paper, EFA 2007 paper.
- Menkveld, A.J., and T. Wang. 2008. “How Do Designated Market Makers Create Value for Small-Cap Stocks?” Working Paper, AFA 2008 paper.
- Mizrach, B., and C.J. Neely. 2006. “The Transition to Electronic Communications Networks in the Secondary Treasury Market.” *Federal Reserve Bank of St. Louis Review* 88:p527–541.
- Naik, N.Y., and P.K. Yadav. 2003. “Risk Management with Derivatives by Dealers and Market Quality in Government Bond Markets.” *Journal of Finance* 58:p1873–1904.
- Nelson, C.R., and A.F. Siegel. 1987. “Parsimonious Modelling of Yield Curves.” *Journal of Business* 60-4:p473–489.
- O’Hara, M. 1995. *Market Microstructure Theory*. Blackwell Publishers.
- Pan, J., and A.M. Poteshman. 2006. “The Information in Option Volume for Future Stock Prices.” *Review of Financial Studies* 19:p871–908.
- Panayides, M.A. 2007. “Affirmative Obligations and Market Making with Inventory.” *Journal of Financial Economics* 86:p513–542.
- Pasquariello, P., and C. Vega. 2007. “Informed and Strategic Order Flow in the Bond Market.” *Review of Financial Studies* 20:p1975–2019.
- Piazzesi, M. 2003. “Affine Term Structure Models.” *Handbook of Financial Econometrics*.
- Poirier, D.J. 1976. *The Econometrics of Structural Change: with Special Emphasis on Spline Functions*. Amsterdam: North-Holland.

- Reiss, P.C., and I. Werner. 1998. "Does Risk Sharing Motivate Interdealer Trading?" *Journal of Finance* 53:p1657–1703.
- Roell, A. 1990. "Dual Capacity Trading and the Quality of the Market." *Journal of Financial Intermediation* 1:p105–124.
- Roll, R. 1984. "A Simple Implicit Measure of the Effective Bid-ask Spread in an Efficient Market." *Journal of Finance* 39:p1127–1139.
- Rosenberg, J.V., and L.G. Traub. 2007. "Price Discovery in the Foreign Currency Futures and Spot Market." Federal Reserve Bank of New York Staff Reports, no. 262.
- Saar, G. 2007. "Systematic Liquidity and Learning about the Risk Premium." Working Paper.
- Stock, J.H., and M.W. Watson. 2007. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit & Banking* 39:p3–33.
- Stoll, H. 2001. "Market Microstructure." Technical Report 16, Owen Graduate School of Management, Vanderbilt University. Prepared for Handbook of the Economics and Finance.
- Stoll, H.R. 1978. "The Supply of Dealer Services in Securities Markets." *Journal of Finance* 33:p1131–1151.
- Stoll, H.R., and R.E. Whaley. 1990. "The Dynamics of Stock Index and Stock Index Futures Returns." *Journal of Financial and Quantitative Analysis* 25:p441–468.
- Van der Wel, M., A.J. Menkveld, and A. Sarkar. 2008. "Are Market Makers Liquidity Suppliers?" Working Paper.
- Vasicek, O. 1977. "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics* 5:p177–188.
- Venkataraman, K., and A.C. Waisburd. 2007. "The Value of a Designated Market Maker." *Journal of Financial and Quantitative Analysis* 42:p735–758.
- Wahal, S. 1997. "Entry, Exit, Market Makers, and the Bid-Ask Spread." *Review of Financial Studies* 10:p871–901.
- Walras, L. 1889. *Elements d'Economie Politique Pure, ou Theorie de la Richesse Sociale*. Rouge.
- Yu, W., and E. Zivot. 2007. "Forecasting the Term Structures of Treasury and Corporate Yields: Dynamic Nelson-Siegel Models Evaluation." Working Paper University of Washington.





The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

383. P. RODENBURG, *The construction of instruments for measuring unemployment.*
384. M.J. VAN DER LEIJ, *The economics of networks: Theory and empirics.*
385. R. VAN DER NOLL, *Essays on internet and information economics.*
386. V. PANCHENKO, *Nonparametric methods in economics and finance: dependence, causality and prediction.*
387. C.A.S.P. SÁ, *Higher education choice in The Netherlands: The economics of where to go.*
388. J. DELFGAAUW, *Wonderful and woeful work: Incentives, selection, turnover, and workers' motivation.*
389. G. DEBREZION, *Railway impacts on real estate prices.*
390. A.V. HARDIYANTO, *Time series studies on Indonesian rupiah/USD rate 1995 - 2005.*
391. M.I.S.H. MUNANDAR, *Essays on economic integration.*
392. K.G. BERDEN, *On technology, uncertainty and economic growth.*
393. G. VAN DE KUILEN, *The economic measurement of psychological risk attitudes.*
394. E.A. MOOI, *Inter-organizational cooperation, conflict, and change.*
395. A. LLENA NOZAL, *On the dynamics of health, work and socioeconomic status.*
396. P.D.E. DINDO, *Bounded rationality and heterogeneity in economic dynamic models.*
397. D.F. SCHRAGER, *Essays on asset liability modeling.*
398. R. HUANG, *Three essays on the effects of banking regulations.*
399. C.M. VAN MOURIK, *Globalisation and the role of financial accounting information in Japan.*
400. S.M.S.N. MAXIMIANO, *Essays in organizational economics.*
401. W. JANSSENS, *Social capital and cooperation: An impact evaluation of a women's empowerment programme in rural India.*
402. J. VAN DER SLUIS, *Successful entrepreneurship and human capital.*
403. S. DOMINGUEZ MARTINEZ, *Decision making with asymmetric information.*

404. H. SUNARTO, *Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of people's credit bank.*
405. M.Â. DOS REIS PORTELA, *Four essays on education, growth and labour economics.*
406. S.S. FICCO, *Essays on imperfect information-processing in economics.*
407. P.J.P.M. VERSIJP, *Advances in the use of stochastic dominance in asset pricing.*
408. M.R. WILDENBEEST, *Consumer search and oligopolistic pricing: A theoretical and empirical inquiry.*
409. E. GUSTAFSSON-WRIGHT, *Baring the threads: Social capital, vulnerability and the well-being of children in Guatemala.*
410. S. YERGOU-WORKU, *Marriage markets and fertility in South Africa with comparisons to Britain and Sweden.*
411. J.F. SLIJKERMAN, *Financial stability in the EU.*
412. W.A. VAN DEN BERG, *Private equity acquisitions.*
413. Y. CHENG, *Selected topics on nonparametric conditional quantiles and risk theory.*
414. M. DE POOTER, *Modeling and forecasting stock return volatility and the term structure of interest rates.*
415. F. RAVAZZOLO, *Forecasting financial time series using model averaging.*
416. M.J.E. KABKI, *Transnationalism, local development and social security: the functioning of support networks in rural Ghana.*
417. M. POPLAWSKI RIBEIRO, *Fiscal policy under rules and restrictions.*
418. S.W. BISSESSUR, *Earnings, quality and earnings management: the role of accounting accruals.*
419. L. RATNOVSKI, *A Random Walk Down the Lombard Street: Essays on Banking.*
420. R.P. NICOLAI, *Maintenance models for systems subject to measurable deterioration.*
421. R.K. ANDADARI, *Local clusters in global value chains, a case study of wood furniture clusters in Central Java (Indonesia).*
422. V. KARTSEVA, *Designing Controls for Network Organizations: A Value-Based Approach.*
423. J. ARTS, *Essays on New Product Adoption and Diffusion.*
424. A. BABUS, *Essays on Networks: Theory and Applications.*
425. M. VAN DER VOORT, *Modelling Credit Derivatives.*
426. G. GARITA, *Financial Market Liberalization and Economic Growth.*
427. E. BEKKERS, *Essays on Firm Heterogeneity and Quality in International Trade.*

428. H. LEAHU, *Measure-Valued Differentiation for Finite Products of Measures: Theory and Applications*.
429. G. BALTUSSEN, *New Insights into Behavioral Finance*.
430. W. VERMEULEN, *Essays on Housing Supply, Land Use Regulation and Regional Labour Markets*.
431. I.S. BUHAI, *Essays on Labour Markets: Worker-Firm Dynamics, Occupational Segregation and Workplace Conditions*.
432. C. ZHOU, *On Extreme Value Statistics*.