CERGE
Center for Economics Research and Graduate Education
Charles University Prague

Essays on Information in Financial Markets

Peter Štefko

Dissertation

Prague, May 4, 2022

Peter Štefko

Essays on Information in Financial Markets

Dissertation

Prague, May 4, 2022

Dissertation Committee

Jan Hanousek (CERGE-EI; chair)

Randall K. Filer (City University of New York)

Jiří Trešl (CERGE-EI)

Referees

George Vachadze (City University of New York)

Eduard Baumöhl (University of Economics in Bratislava)

To my parents, for all their patience and encouragement.

Table of Contents

Abstract

The first chapter of this dissertation carries out a multidimensional investigation of weak-form market efficiency for the stock market indices of the United States, Germany, and five CEE countries during the years following the 2008 Global Financial Crisis. We analyze developments over time in the predictable power and potential profitability of three classes of well known technical trading rules applied to market index prices, sampled at three different frequencies: daily, hourly, and 15-minute. We approach our analysis by testing for true abnormal performance of trading rules using White's Reality Check procedure as well as by evaluating a simple out of sample investment strategy based on ex-post best performing trading rules. We find that, while the developed stock markets of the US and Germany exhibit results strongly consistent with market efficiency at all three sampling frequencies, some evidence of predictable power of technical analysis (consistent with the concept of adaptive efficiency introduced by Lo (2005)) is present for the stock markets of Austria, the Czech Republic, Hungary, Poland, and Slovenia at the two intraday sampling frequencies. However, the results are challenged when we take transaction costs into account. Our proposed comparison of developed and emerging stock markets provides new insights regarding the degree of market efficiency, especially for intraday data and for countries that have experienced a transition to a market economy relatively recently.

The second chapter utilizes trade-level data on WTI crude oil futures contracts during the period 2017 - 2019 to assess the dynamics of market microstructure order flow toxicity around the weekly regular US crude oil inventory levels announcements. We estimate order flow toxicity using the Volume-synchronized Probability of Informed Trading (VPIN) metric devised by Easley, de Prado, and O'Hara (2012) and we investigate announcements by two different publishers. We find that this probability metric increases on average by 20% immediately (within seconds) after data publication and the magnitude of this change is positively associated with the degree to which the particular announcement can be regarded as surprising. Moreover, for the report published by the US Department of Energy, we observe 20% smaller average values of daily VPIN profile during the 30 minutes leading up to the publication time, as opposed to the same time of day on non-announcement days.

The third chapter investigates risk contagion dynamics within a set of the four most liquid Eurozone sovereign government bond futures (France, Germany, Italy, and Spain) during the period 2016 - 2020, applying the Diebold and Yilmaz (2014) connectedness

framework. We find that the overall system-wide connectedness, as measured by the Total Spillover Index, varies between 30 and 72 percent, representing a moderately interconnected system when compared to other benchmark studies. The connectedness exhibits the tendency to rise sharply following geopolitical or macroeconomic shocks, represented in our sample mainly by the Italian political crisis of 2018 and the Covid-19 outbreak of 2020. Italian government bond futures are the principal spillovers transmitter vis-à-vis the rest of the system, despite the fact that the Italian bond market accounts for a proportionately smaller fraction of the system's aggregate traded volume. While the volatility transmitting role of Italy is predominantly associated with its internal turbulent political development, Germany takes the leading role as the sample's volatility transmitter of major global affairs. We additionally find evidence of a significant association between the degree of a country's directional spillovers transmission and its underlying macroeconomic fundamentals, particularly the changes in the country's debt-to-GDP ratio.

Abstrakt

První kapitola této disertační práce provádí vícerozměrnou analýzu slabé formy efektivnosti trhu pro akciové indexy Spojených Států, Německa a pro pět států střední a východní Evropy v období následujícím Světovou finanční krizi roku 2008. Analyzujeme časový vývoj predikční schopnosti a potenciální profitability třech tříd populárních pravidel technického obchodování aplikovaných na ceny trhových indexů, uvažujíc tři různé frekvence vzorkování cen: denní, hodinovou a 15 minutovou. K analýze přistupujeme testováním skutečné mimořádné výkonnosti technických pravidel použitím Whitovy Reality Check procedury a vyhodnocením jednoduché investiční strategie založené na ex-post nejvýkonnějším obchodovacím pravidle. Zjišťujeme, že zatím co rozvinuté akciové trhy Spojených Států a Německa vykazují výsledky silně konzistentní s trhovou efektivností na všech vzorkovacích frekvencích, přítomnost predikční schopnosti technické analýzy (konzistentní s konceptem adaptivní efektivností uvedené v Lo (2005)) je pozorována pro akciové trhy Rakouska, České Republiky, Maďarska, Polska a Slovinska na dvou vnitrodenních frekvencích. Výsledky jsou avšak zpochybněny v případě zohlednění možných transakčních nákladů. Naše porovnání rozvinutých a rozvíjejících se akciových trhů přináší nové výsledky ohledně rozsahu tržní efektivnosti, obzvlášť pro vnitrodenní data a pro státy, které absolvovali transformaci na tržní ekonomiku pouze relativně nedávno.

Ve druhé kapitole používáme data na úrovni jednotlivých obchodů k popisu dynamiky toxicity objednávkového toku v tržní mikrostruktuře ropných WTI futures kontraktů v okolí pravidelných týdenních oznámení úrovní amerických ropných zásob v průběhu 2017 - 2019. Toxicitu objednávkového toku odhadujeme pomocí VPIN pravděpodobnostní metriky vytvořené v Easley, de Prado a O'Hara (2012) a zkoumáme oznámení dvou rozličných agentur. Zjišťujeme, že tato pravděpodobnostní metrika narůstá v průměru o 20% okamžitě (řádově v rozmezí sekund) po publikaci dat a velikost této změny je pozitivně asociována s mírou překvapení daného oznámení. Pro oznámení publikované americkým ministerstvem energetiky navíc pozorujeme o 20% nižší průměrné hodnoty denního VPIN profilu během 30 minutového okna vedoucího k času oznámení, v porovnání se stejnou částí dne ve dnech bez oznámení.

Ve třetí kapitole zkoumáme dynamiku přelévání tržního rizika v rámci skupiny čtyř nejlikvidnějších evropských futures kontraktů na státní dluhopisy (Francie, Německo, Itálie a Španělsko) v období 2016 - 2020, aplikujíc rámec metriky propojenosti uveden v Diebold a Yilmaz (2014). Zjišťujeme, že celková úroveň propojenosti systému, měřená metrikou Total Spillover Index, se pohybuje v rozmezí 30 až 72 procent,

představujíc středně propojen systém v kontextu porovnatelných studií. Propojenost vykazuje tendenci prudkých nárůstů v období po geopolitických a makroekonomických otřesech, které jsou v naší studii zastoupeny italskou politickou krizí v roku 2018 a vypuknutím pandemie Covidu-19 v roku 2020. Italské futures státních dluhopisů jsou hlavním původcem rizika ve vztahu ke zbytku systému, navzdory tomu, že italský trh státních dluhopisů reprezentuje relativně malý zlomek celkového zobchodovaného bohatství v rámci námi uvažovaného systému. Zatímco role původce volatility Itálie je spjatá především s jejím vnitřním turbulentním politickým vývojem, Německo v našem vzorku přebírá vedoucí roli původce volatility způsobené vývojem celosvětovým. Navíc nacházíme důkazy o významné asociaci mezi mírou směrového přelévání rizika země a jejími základními makroekonomickými fundamenty, zejména změnami v poměru dluhu dané země k HDP.

Acknowledgments

I would like to express my utmost gratitude to my supervisor, Jan Hanousek, not only for all his exceptionally valuable and professional scientific advice, but also for his overall research guidance, support, and practical insights I drew upon throughout the entire duration of my dissertation research. His cordial attitude and endless patience made my studies much more manageable.

I am also grateful to the members of my dissertation committee, Randall K. Filer and Jiří Trešl, and to my referees, Eduard Baumöhl and George Vachadze, for dedicating their time and effort in order to provide me with extraordinarily useful feedback that led to a further improvement of the quality of my thesis.

Additionally, I am highly grateful to Deborah Nováková for the detailed and very valuable editing of the final text, and to the CERGE-EI SAO staff members, for all their help over the years.

Finally, I warmly thank my parents for their limitless support, for all their words of encouragement, and for instilling in me, as well as in my brothers, an appreciation for the value of education.

Financial support received from Charles University in Prague (GAUK project No. 470217) is gratefully acknowledged.

Czech Republic, Prague
May 4, 2022

Peter Štefko

Introduction

As the central theme intertwining each of its three chapters, this dissertation thesis presents empirical investigations of various qualitative aspects of information dissemination dynamics in financial markets. Each of the three studies examines different asset classes, different time horizons and frequency granularities, and different varieties and manifestations of information arrival which I assume affect the asset valuations under consideration.

The first chapter investigates perhaps the most popular (and the most contradictory of mainstream economic and asset pricing theory) practice of non-professional investing, the practice of exploiting past prices to forecast future prices, also known as technical trading. Under certain (admittedly not implausible) assumptions, the Efficient Markets Hypothesis (EMH, originally formulated by Fama (1970)) asserts that the returns on asset prices ought to be unpredictable, that is, it should be impossible to develop a forecasting model based on past price and volume history data (the so-called technical trading rule) that would consistently generate positive risk-adjusted expected excess returns.

The key word in the above sentence is "consistently". Indeed, because, given a particular time series of prices, there is a virtually infinite number of possible trading rules one could devise; by trying long enough, one is certain to find at least one trading rule that outperforms any given stochastic process ex-post, even if the process is merely a random sequence of numbers, and the trading rule does not possess any true predictive power. Therefore, in order to properly assess the validity of market efficiency, suitable statistical methods accounting for the potential presence of data-snooping bias need to be employed.

Acknowledging these considerations, the first chapter analyzes the performance of a set of 616 mainstream technical trading rules applied to stock market indices of the United States, Germany, and five Central and Eastern European countries during the years following the 2008

Great Financial Crisis. We add a dimension to the results by considering three different sampling frequencies (daily, hourly, and 15-minute), under which signals generated by the trading rules can be acted upon. We use the White Reality Check procedure, and an out-of-sample investment performance evaluation of in-sample best-performing trading rules to control for the possibility of data-snooping bias.

We find that, while the developed stock markets of the United States and Germany exhibit results strongly consistent with market efficiency at all three sampling frequencies, some evidence of the predictable power of technical analysis is present for the stock markets of Austria, the Czech Republic, Hungary, Poland, and Slovenia at the two intraday sampling frequencies. Although the results are challenged when transaction costs are taken into account, certain qualitative differences regarding the degree of market efficiency between developed and emerging stock markets are observed, especially for intraday data and for the countries that have relatively recently transitioned into market economies.

In the second chapter, we analyze the information adaptivity of price discovery processes from a high-frequency market microstructure perspective. Trade-level data on WTI crude oil futures contracts are used to investigate the dynamics of order flow toxicity, one of the central concepts of the academic literature on market microstructure. One of the distinct features of this market is the presence of regular weekly macroeconomic data announcements with significant impacts on prices: the publications of the changes in the levels of US crude oil inventories. We consider two different reports: the Weekly Statistical Bulletin by the American Petroleum Institute (API) and the Weekly Petroleum Status Report by the Energy Information Association of the US Department of Energy (DOE).

To assess the manner in which the announced inventory information dissipates in the price discovery process, we apply the VPIN metric by Easley, de Prado, and O'Hara (2012), a non-parametric procedure allowing us (in combination with the extremely low granularity of our data) to identify the changes in order flow toxicity at relatively high frequencies. The VPIN metric quantifies the order flow toxicity by the estimated probability of informed trade arrival.

We document that the estimated probability of informed trading increases by approximately 20% on average for both types of reports under investigation. Though in both cases it occurs virtually immediately, the most pronounced VPIN adjustment after publication of a DOE report takes place, on average, within the first two seconds, about three times faster than after the API report. Additionally, we find some evidence of a positive association between the magnitude of the order flow toxicity change around the two types of inventory announcements and the degree to which these announcements are surprising. Interestingly, on DOE announcement days, we observe, on average, 20% smaller values of the daily VPIN profile during the 30 minutes leading up to the publication time than on non-announcement days at the same time of day. We conjecture that this phenomenon may be associated with relatively greater activity of liquidity

demanding traders aiming to reduce their oil price exposure due to the upcoming announcement, as well as with relatively lower activity of informed and speculative traders, who are less willing to engage in their usual trading activity. This observation may serve as an example of potentially valuable microstructure-related nuances that can be uncovered by applying the VPIN metric to electronic limit order book market data.

Finally, we consider the dynamics of information dissemination interactions within a set of multiple assets in the third chapter of the thesis. We use the connectedness framework of Diebold and Yılmaz (2014) to evaluate the magnitude and directions of risk contagion within the set of the four most liquid Eurozone sovereign bond futures (pertaining to France, Germany, Italy, and Spain) during 2016 - 2020, a period including multiple geopolitical and macroeconomic events of substantial regional and global significance, including the adoption of protectionist economic policies by the Trump administration, the Italian political crisis, and the Covid-19 pandemic outbreak.

We find that the Eurozone sovereign bond market, as represented by our sample, constitutes a system entailing considerable, although not extreme (compared to other asset classes investigated in academic literature) degree of risk contagion, ranging between 53 and 60 percent, as measured by the overall system connectedness indicator (the so-called Total Spillover Index). When we consider the entire sample, we observe the greatest directional contagion effect for volatility spillovers in the "from Italy to others" direction. We argue that the turbulent political situation in Italy present throughout a substantial time period of the sample was associated not only with greater volatility of Italian government bond futures, but also measurably contributed to increased volatility of the bond markets of the remaining major Eurozone countries. We find this particularly interesting, as the Italian bond futures consistently generated only around 10% of the aggregate system's traded volume.

Additionally, we observe that particular aspects of the system's connectedness exhibit significant time variation. The total connectedness of the system, for example, tends to increase rapidly in a response to turbulent and unexpected macroeconomic or geopolitical shocks. Finally, though we find that the volatility transmitting role of Italy has been mainly caused by its internal turbulent political development, Germany is the main volatility transmitter of major global affairs in our sample. The coronavirus outbreak market-wide volatility surge (ignited on March 12, 2020) represents an event for which this dynamics is observed in the most pronounced manner.

Chapter 1

Does the Profitability of Technical Analysis Depend on Trading Frequency? A Comparison of Developed and Emerging Markets.

1.1 Introduction

How rapidly do stock prices change and adapt to new information? Is it possible to forecast future price development? The phenomena of price discovery in financial markets and its possible predictabililty and exploitability with the aim of making economic profits have for long been receiving considerable attention of both market practitioners as well as academicians.

In the academic sphere, the concept that has been strongly dominant is the Efficient Markets Hypothesis (EMH), originally postulated by Fama (1970). In order to establish the definition of market efficiency, the following extended formalization by Jensen (1978) is useful: "A market is efficient with respect to information set θ_t if it is impossible to make economic profits by trading on the basis of information set θ_t." Therefore, the prices at time t in an efficient market fully incorporate the content of θ_t, such that it is impossible to use the information content of θ_t to obtain risk-adjusted net returns by forecasting future prices. Effectively, everything that could be learned from θ_t has already been incorporated in the price at time t. Jensen's taxonomy also uses the three basic types of market efficiency as originally proposed by Fama (1970):

- **Weak form efficiency:** θ_t contains only the history of financial market data, such as stock prices, volume, and open interest available at time t.
- **Semi-strong form efficiency:** θ_t contains all publicly available information available at time t.

- **Strong form efficiency:** θ_t contains both public and private information available at time t.

One particular implication of market efficiency is that it is impossible to develop a forecasting model based on price and volume history (the so-called *technical trading rule*) that would generate positive risk-adjusted expected excess returns. However, the activity of developing such models, known also as the *technical analysis*, has been extremely popular in the financial sector. For example, Cheung and Chinn (2001) document that 30% of US foreign exchange traders may be characterized as technical traders, with the proportion exhibiting an upward trend. Popularity of technical analysis in financial industry may be illustrated by the large number of textbooks (for instance, Murphy (1999), Grimes (2012) and Pring (2021)) and popular investment websites (such as `https://www.seekingalpha.com`, `https://www.marketwatch.com` and `https://www.investopedia.com`) related to technical analysis. Technical analysis is being applied in financial markets and it is being applied to a significant extent. Its relevance and rationale in the investor's portfolio selection process is therefore interesting also from the academic perspective.

Despite the dominant position of the EMH in the economic theory during the twentieth century, by the 1990s, its general acceptance by academic community began to diminish. Jensen (1978) asserts that EMH is the most vigorously empirically supported hypothesis in economics. On the other hand, Grossman and Stiglitz (1980) go even as far as to claim that the perfect efficiency of financial markets is an impossibility, since under the assumption of EMH, no incentives to gather information are present. With the onset of behavioural finance (motivated especially by the experimentally documented irrationality of decision making by Kahneman and Tversky (2013) and by its manifestations in financial markets by De Bondt and Thaler (1985)), as well as the increasing number of empirical studies yielding conflicting results (see a survey of the EMH history by Sewell (2011)), alternative hypotheses began to emerge. Campbell, Lo, and MacKinlay (2012) introduced the concept of relative efficiency, no longer perceiving the efficiency of markets as an absolute characteristic. Eventually, inspired by concepts in evolutionary biology and neuroscience, such as the natural selection, competition and mutation, Lo (2005) proposed the Adaptive Markets Hypothesis (AMH), claiming that the degree of market efficiency is highly context-dependent and evolving over time. According to the AMH, any examples of irrationality, such as overconfidence or overreaction are assumed to be consistent with gradual, iterative adaptation of market participants to the constantly changing market environment and this process is likely to offer potentially profitable (albeit short lived) market inefficiencies.

While the profitability of technical analysis is in direct contradiction with the EMH, it does not necessarily violate its alternatives. Therefore, not surprisingly, another branch of academic research that contributed to the (still ongoing) debate regarding the validity of the EMH tests for the presence of technical trading rules (TTRs) that might contain true predictive power

over the stock price process. After the seminal work of Brock, Lakonishok, and LeBaron (1992) that provided strong evidence of predictable power of simple TTRs applied to the DJIA index, the popularity of academic literature devoted to technical analysis increased rapidly. In their excellent survey of this topic, Park and Irwin (2007) note that out of 95 studies dated after 1988 that they consider, 56 studies provide evidence against the EMH, while 20 studies find that technical analysis does not lead to excess returns. However, the authors note that there is a large number of methodological issues connected with testing for the predictable power of technical analysis and the approach in which these issues are considered is not consistent across different studies.

Among such issues are the proper risk adjustment of the trading strategy returns (since a possible higher expected return may simply be a compensation for higher risk) or accounting for transaction costs (since even if a trading rule possesses predictive power over the price process, in order to obtain economic profits, one needs to be able to trade the given asset as often as the rule requires, while simultaneously bearing transaction costs that do not wipe out all its excess returns).

Additionally, as in any other empirical analysis, statistical significance of the results needs to be addressed properly. In order to investigate market efficiency, technical analysis studies usually analyze large numbers of TTRs. Indeed, there is virtually an infinite number of possible trading rules one could devise and apply to data. By trying long enough, one is certain to find at least one trading rule that outperforms any given stochastic process *ex-post*, even if such process is a purely random sequence of numbers and the trading rule does not bear any true predictive power. When multiple models are applied to the same dataset, the possibility of the *data snooping bias* in the results arises.

In this study, we contribute to the market efficiency related debate by evaluating the predictable power of 616 technical trading rules applied to seven different stock market indices observed during the time period 2008 - 2015. We are mainly interested in assessing whether the EMH or the AMH is a better descriptor of the reality. In this respect, we contribute in several dimensions.

Fistly, we note that while the other studies focus on a wide variety of different markets (such as developed and emerging stock markets, foreign exchange markets, derivatives or commodity markets), the emerging markets of the Central and Eastern Europe (CEE) have not been analyzed sufficiently. We consider five CEE countries, namely, the stock market indices of Austria, the Czech Republic, Hungary, Poland, and Slovenia. Moreover, the results for the CEE region are qualitatively compared to the well-developed and highly liquid stock markets of the United States and Germany. This sample of stock market indices represents an interesting collection of both developed as well as emerging financial markets that vary in terms of liquidity or market microstructure characteristics.

Secondly, while the vast majority of studies usually focus on daily, weekly or even monthly data, our study utilizes three different sampling frequencies: daily, hourly, and 15-minute. We therefore consider also intraday sampling frequencies, a feature less often present in technical trading studies. The main motivation is our hypothesis that even if trading rules prove to be not useful for trading at lower frequencies (such as on a daily basis, a frequency trading at which may be pursued by virtually any investor, regardless of their institutional background), an investor who has the possibility of trading at higher (intraday) frequencies may still be able to uncover short-lived market inefficiencies that can possibly be exploitable by technical analysis. Indeed, as Marshall, Cahan, and Cahan (2008) note, surveys reveal that the weight placed on technical analysis in trading at intraday frequencies almost doubles as opposed to longer horizons and smaller frequencies. By studing intraday data and three different frequencies, we are able to evaluate market efficiency not only across markets, but also to study its dependence on the sampling frequency of the price process.

Thirdly, besides controlling for the presence of the data snooping bias by the means of the White (2000) Reality Check procedure, the structure of our analysis also enables us to conduct a simple out of sample predictability evaluation of the trading rules considered in this study. In particular, we assess how an investor identifying the *ex-post* best performing trading rules on a regular basis and subsequently following them in the market (the so-called "Best-Rule" strategy) performs as opposed to an investor that simply follows the market by passively holding the market index. Moreover, we also provide a discussion on whether transaction costs are likely to affect the potential profitability of this strategy.

Finally, because of the non-overlapping subsamples approach this study undertakes, the development of all results across time might be observed. This feature also contributes to the market efficiency debate, since under the EMH, technical analysis should not be profitable in any of the subsamples, while under the AMH, market inefficiencies may arise temporarily.

Our results indicate that our sample of seven stock market indices may be objectively divided into two groups. The first group, consisting of the S&P 500 and the German DAX, exhibits results that strongly support market efficiency. For the vast majority of subsamples, White's Reality Check fails to reject the null hypothesis of no superior performance of the best rule. Moreover, using the best performing rules in the "Best-Rule" investment strategy does not yield excess performance significantly greater than zero. These results are stable over time and robust with respect to the choice of the performance measure or the sampling frequency of the price process.

The second group of indices, consisting of the five CEE indices exhibits similar results only for the daily frequency data. For the two intraday frequencies, we find some evidence of time varying informational efficiency of the markets with respect to our set of trading rules. Moreover, an investor following the "Best-Rule" investment strategy would be able to expect excess

performance that is greater than zero in a statistically significant manner for three indices at the hourly frequency and for all five indices at the 15-minute frequency. This excess performance varies from 13% p.a. for the Slovenian index and the daily frequency trading to as much as 73.2% p.a. for the Austrian index and the 15-minute frequency, thus offering potentially very interesting returns before transaction costs.

Overall, we argue that moving from well-developed towards emerging stock markets, as well as moving from lower to higher intraday frequencies is associated with results that challenge the concept of market efficiency to a greater degree. Our analysis contributes to the informational efficiency-related debate that is of great importance, especially for the financial markets that may still be considered emerging.

The remainder of this paper is organized as follows. Section 1.2 provides a review of studies that share a common ground with our study. Section 1.3 describes all the methodological aspects we apply, while section 1.4 introduces the data in a detailed manner. Section 1.5 provides the results, that we divide into two main parts: the White's Reality Check resuls and the "Best-Rule" investment strategy evaluation. Section 1.6 concludes.

1.2 Literature Review

1.2.1 Overview of the Main Studies

Empirical studies devoted to technical analysis have been around for decades. Early works include Fama and Blume (1966), Jensen and Bennington (1970) and Kaufman (2013). Understandably, these early studies focus mostly on developed American stock market indices sampled at relatively low frequency (such as monthly, weekly or daily).

In their seminal study, Brock, Lakonishok, and LeBaron (1992) (BLL) analyze two of the most common trading rules, namely, the Moving Average (MA) and the Trading Range Breakout (TRB). Utilizing over 90 years worth of daily data on the DJIA index and 26 inividual trading rules, the authors document strong evidence of profitability. The authors are also the first to apply the model-based bootstrap methodology to assess the statistical significance of the results[1] and conclude that each of the 26 trading rules generate returns that outperform the market.

Another study providing evidence of the TTR profitability is Levich and Thomas III (1993).

[1]The process consists of assuming a null stochastic process for the stock price, estimating its parameters and randomly resampling the residuals with replacement. BLL use four different assumptions for the null process: a random walk with drift, an AR(1), a GARCH-M and EGARCH. By performing 500 bootstrap resamplings, the authors estimate the underlying probability distribution of the trading rules' returns under each of the null models. This procedure overcomes the usual t-tests, which assume independent and indentically distributed data, an assumption that is usually violated for asset returns. However, inconsistent results may be obtained by misspecifying the null models, as noted by Maddala and Li (1996).

Focusing on the FX market and analyzing 15 years worth of FX-futures data, the authors report a large number (60 %) of statistically significant profitable TTRs. These results are robust with respect to 5-year subperiods, even though a moderate decline in profitability is reported for the last subperiod.

The BLL study is one of the most important in the area of technical analysis. Because of their significant results, the intensity of the subsequent research increased rapidly. Bessembinder and Chan (1998) applied the BLL set of trading rules to the DJIA index adjusted for dividends and conclude that after transaction costs, the BLL rules do not lead to profitable strategies.

Sullivan, Timmermann, and White (1999) (STW) are the first to apply the White (2000) Reality Check procedure designed to test for the presence of data snooping. The authors develop a "universe of trading rules", containing over 7000 individual TTRs and acknowledge that under such setting, the dangers of data snooping are strongly present. The Reality Check tests the null hypothesis that the best performing trading rule does not bear predictive power over the market. Applying the procedure to the DJIA index over 1987 - 1996, as well as to the S&P 500 index futures over 1984 to 1996, the authors report that while the trading rules were indeed profitable in the original BLL dataset (also after the data-snooping is controlled for), these properties diminished in the additional 10 years of data (i.e., out of sample of the BLL study). The Reality Check procedure has gained major attention (not only) in the subsequent technical analysis related studies.

Other studies focus on more complex technical trading algorithms. Gencay (1999) applies the nearest-neighbor and the feed-forward neural network algorithms to five currency pairs and finds that these algorithms are capable of significantly outperforming the market benchmark. Neely and Weller (2003) use genetic programming to find profitable trading rules at intraday foreign exchange market data and conclude that the optimal trading rules do not outperform the market. Recently, Gradojevic and Gençay (2013) analyze the performance of fuzzy-logic algorithms and find that these algorithms outperform standard moving average rules for the Euro - US Dollar currency pair, especially during periods of higher volatility.

Bajgrowicz and Scaillet (2012) continue with the analysis of the daily DJIA index data and the universe of trading rules developed by Sullivan, Timmermann, and White (1999). The authors focus on mathematically rigorous approach towards the issues of risk-adjustment, data-snooping as well as transaction costs. The profitability of TTRs is seriously questioned, especially during the last two decades.

Since its inception by Lo (2005), numerous studies specifically analyzed the extent to which empirical observations conform to the Adaptive Market Hypothesis (AMH). The literature that provides evidence on time-varying stock market predictability is expanding (Lim and Brooks (2011)). Neely, Weller, and Ulrich (2009) compare the ability of EMH and AMH in explaining foreign exchange market conditions and find evidence in favor of the AMH. Kim, Shamsuddin,

and Lim (2011) provide another evidence in favor of the AMH. Applying the variance ratio as well as automatic portmanteau tests to both daily and weekly DJIA data, the authors conclude that the degree of market efficiency substantially varies over time and is governed by dynamic market conditions. Fundamental financial, economic and political crises significantly change the efficiency of financial markets. Zhou and Lee (2013) argue that the AMH provides suitable framework for the explanation of the US real estate market. Urquhart and McGroarty (2014) analyze various calendar anomalies and conclude that empirical results based on daily DJIA data are strongly supportive of the AMH.

1.2.2 Evidence of Emerging Markets

One of the first studies analyzing technical trading in emerging markets is Bessembinder and Chan (1995). Considering daily data for three developed (Hong Kong, Japan and South Korea) and three emerging (Malaysia, Taiwan and Thailand) markets, the authors observe that while simple trading rules are not useful in the developed markets, there is evidence of their predictable power over the three emerging markets. Lee, Pan, and Liu (2001) apply similar trading rules to 13 daily currency spot rates in Latin America and conclude that the rules are able to predict movements especially for the currency pairs of Brazil, Mexico and Venezuela. Similar conclusion for the Brazilian foreign exchange spot rate is also obtained by Tabak and Lima (2009). In particular, the moving average and trading range breakout rules appear to have predictive ability during the period 1999 - 2006. However, the authors state that transaction costs are likely to wipe out all the potential excess returns.

Another emerging market that exhibits certain characteristics suggesting it may be inefficient, namely, the equity market of New Zealand, is analyzed by Marshall and Cahan (2005). However, applying 12 popular trading rules, the authors do not find evidence of their profitability. Stock markets of Brazil, China, India and Russia are studied by Chong, Cheng, and Wong (2010). It is documented that simple technical indicators are most profitable in the Russian stock market, while the Brazilian stock market is found to be the most efficient in this sample.

Yu et al. (2013) focus on four developing and one developed Asian stock markets and analyze the profitability of the MA and TRB rules during the period 1991 - 2008. These technical rules are documented to outperform the market for developing stock markets. Nevertheless, by accounting for transaction costs, economic significance of this profitability diminishes. The authors highlight the need of constant verification of statements regarding market efficiency, especially for emerging and growing stock markets. Among other studies that focus on emerging equity markets and obtain some evidence of profitability we mention the analyses of the African markets by Ntim et al. (2011), Australian market by Park et al. (2014), the individual stocks of the Taiwanese market by Ko et al. (2014) and the Ugandan stock market by Katusiime,

Shamsuddin, and Agbola (2015).

Despite this relatively varied mixture of emerging stock markets and foreign exchange rates of emerging countries, surprisingly low attention has been devoted to the technical trading profitability in emerging markets of the Central and Eastern Europe. Fifield, Power, and Donald Sinclair (2005) consider 11 European stock markets during the 1990s and report that while the developed markets are informationally efficient, the emerging markets considered are inefficient, even after transaction costs are accounted for. Moreover, the authors find that while the moving average rules often do not appear to be successful, filter rules consistently outperform the market.

In their extensive study, Marshall, Cahan, and Cahan (2010) use the MSCI data of 49 countries, including Austria, Czech Republic, Hungary and Poland to study 5806 popular TTRs. While some qualitative differences between emerging and developed markets are observed, the authors note that these differences are not particularly strong after data snooping is controlled for, when daily data are considered. Metghalchi, Marcucci, and Chang (2012), however, report that moving average rules manage to outperform some of the 16 European stock markets between 1990 and 2006 on a daily basis, even after the White (2000) Reality Check is applied to account for data snooping. Finally, Heyman, Inghelbrecht, and Pauwels (2012) utilize a large daily frequency sample of 34 stock market indices, including the Czech Republic, Hungary and Poland. After transactions costs as well as the effects of data snooping are controlled for, the authors report significant profitable trading rules only for market indices of Botswana, Jamaica, Kenya and Oman.

1.2.3 Evidence of Intraday Frequencies

As mentioned earlier, the majority of studies have utilized price data sampled at relatively low frequencies. The most frequently used sampling frequency is daily, while weekly or even monthly frequencies were not uncommon in earlier studies. The number of studies utilizing intraday frequency data remains rather modest. As Park and Irwin (2004) note, out of 92 considered studies, only 6 make use of intraday data and all 6 of them focus on currency pairs.

For example, Osler (2000) investigates the support and resistance levels for multiple currency pairs sampled at 1-minute frequency and reports that the specific support and resistance levels bear certain predictive power. Neely and Weller (2003) use genetic programming to devise trading rules on four major currency pairs data sampled at 30-minute frequency. For all four currencies, the authors document predictable patterns in the intraday data. This predictability is found to be exceptionally stable, however, once transaction costs are controlled for, the presence of excess returns for the devised trading rules disappears.

More recently, several authors have focused also on intraday frequencies in equity markets.

Marshall, Cahan, and Cahan (2008) focus on US equity market approximated by the S&P Depository Receipts data at 5-minute frequency. By studying more than 7000 simple trading rules, the authors conclude that even though a small number of rules appears to be profitable, once accounting for data snooping bias, any profitabililty diminishes. Schulmeister (2009) analyzes spot and futures S&P 500 market and observes qualitative differences in market efficiency between daily and 30-minute frequency data. In particular, between 1983 and 2000, technical analysis applied on 30-minute frequency yields on average 8.8% of gross return, while the profitability at daily frequency gradually disappears during the same period. Finally, Yamamoto (2012) uses 5-minute data on Japanese equity markets and concludes that technical analysis applied in such setting is not profitable.

Recently, Frömmel and Lampaert (2016) examine the profitability of two classes of TTRs (the Moving Average and the Relative Strength Indicator) applied on the US Dollar - Russian Ruble currency pair sampled at 60, 30, 15 and 10 minute frequencies. The authors conclude that despite the fact that technical analysis seems to possess greater predictive power at higher frequencies, the profitability of technical trading is highly questionable after accounting for transaction costs.

To the best of our knowledge, there does not currently exist a study that would analyze intraday frequency data of CEE equity markets.

1.3 Methodology

1.3.1 Behavioural Justification for Technical Analysis

Why would technical analysis work? The following definition of technical analysis due to Pring (2021) is illustrative in answering this question:

"The technical approach to investment is essentially a reflection of the idea that prices move in trends that are determined by the changing attitudes of investors toward a variety of economic, monetary, political, and psychological forces. The art of technical analysis, for it is an art, is to identify a trend reversal at a relatively early stage and ride on that trend until the weight of the evidence shows or proves that the trend has reversed."

Indeed, the common objective of virtually every possible TTR is to identify the direction of trend in the future price development based on past (and current) information. If, according to the particular TTR, the price is likely to trend upwards, or if the downward trend is likely to be reversed in the near future, it is advantageous to be long in a security, therefore, the TTR provides the *buy signal*. Conversely, if downward trend is predicted, the *sell signal* is provided and entering a short position is advised. Finally, if there is not enough evidence for a trend, the *neutral signal* may be present. One of the main justifications for the (hypothesized) trending

behaviour of prices is that new information is incorporated into prices in a gradual manner, while being exposed to various accomodative tendencies exhibited by market participants. If such complicated price discovery arises from behavioural tendencies (such as herding, overreaction or overconfidence, aspects of human decision making documented in great detail by behavioural finance literature[2]), exploitability of systematic patterns in such behaviour by analyzing past data becomes possible.

The characteristics outlined above provide some justification for heavy employment of technical analysis in practice.[3] Therefore, there might be a strong belief among market practitioners that analyzing past prices is important, whether it is because it contains true economic information or simply because they think that other investors possess these beliefs and a risk of self-fulfilling prophecies is present.[4] Technical trading rules do not just appear out of nowhere. Instead, they represent a long lasting manifestation of the assumptions some investors make about the price discovery processes. If markets are informationally efficient, these assumptions are not valid and technical analysis does not consistently yield risk-adjusted returns that outperform the market. It is this market-outperforming power of technical analysis that we aim to examine in this study.

1.3.2 Performance Evaluation

Before we introduce the particular set of TTRs used in this study, we need to establish the framework in which the performance of the TTRs will be evaluated. An approach established in literature is to calculate certain function of the returns that the trading strategy provides and compare the value to the returns provided by some benchmark strategy.

Assume that by $t \in \{1, 2, \ldots, T\}$, we denote the time indices corresponding to the sample during which a trading rule $k \in \{1, 2, \ldots, K\}$, where $K = 616$ is the total number of TTRs, is applied and evaluated. As already discussed, the ultimate output of each trading rule is a trade recommendation by signaling the desired position in an asset. A trading rule k thus provides an investor at each time period t with a signal $S_t^{(k)}$, where the interpretation of the signal is as follows:

$$S_t^{(k)} = \begin{cases} 1 & \text{during time period } t+1, \text{ hold a long position} \\ 0 & \text{during time period } t+1, \text{ hold a neutral position} \\ -1 & \text{during time period } t+1, \text{ hold a short position} \end{cases}$$

Hence, we assume that upon a trading signal at time t, an investor follows this signal by

[2]Excellent surveys of behavioural finance literature are provided by Barberis and Thaler (2005) and De Bondt et al. (2008).

[3]A very detailed discussion of both the theoretical and empirical justifications for why technical analysis might be profitable is provided by Park and Irwin (2007).

[4]Self-fulfilling prophecy argument is also made by Froot, Scharfstein, and Stein (1992).

executing the corresponding trade (buy/sell/do nothing) at a closing price P_t and the corresponding one-period log-return emanating from this signal is based on prices P_t and P_{t+1}. Since at the moment of generation of $S_t^{(k)}$, the price P_t is not yet known, the signal must be based on the information available up until (and including) time $t-1$. The signal $S_t^{(k)}$ is not a function of P_t in our study, since we wish to account for the possibility of non-synchronous trading (when an investor is not able to execute the trade at the price based on which the trade decision is made), that may pose a real problem especially for trading at higher frequencies.

Since market efficiency precludes the existence of trading rules that would be able to consistently outperform the market, we evaluate each trading rule against the buy-and-hold benchmark strategy consisting of entering a long position at time $t = 1$ and holding this position until $t = T$. Since a long position is assumed to be the default benchmark position, our trading rule signals will superimpose a position on this buy-and-hold strategy.[5] Thus, a long position generated by a buy signal may be interpreted as doubling the investment in an asset by borrowing the resources needed to leverage the position, since we have a good reason to believe that the price will go up. A neutral signal may be interpreted as simply holding an unlevered position in an asset (equivalent to the buy-and-hold strategy) and a short position emanating from a sell signal may be interpreted as liquidating the position in an asset and staying out of the market, since we have a good reason to believe that the price will decline. We also assume that in order to leverage, investor borrows the required funds at a (continuously compounded) risk-free rate of return $r_t^{(F)}$. This rate of return is also earned by investor while staying out of the market.

Denote the one-period time t continuously compounded rate of return of the market as $r_t^{(BH)}$, where:

$$r_t^{(BH)} = \log\left(\frac{P_t}{P_{t-1}}\right). \tag{1.1}$$

The return $r_t^{(BH)}$ is equivalent to the return of the buy-and-hold strategy, since this strategy simply follows the market. The corresponding time t gross log return of the trading rule k at time t is (according to our description above) given by:

$$r_t^{(k)} = \begin{cases} 2r_t^{(BH)} - r_t^{(F)} & S_{t-1}^{(k)} = 1 \\ r_t^{(BH)} & S_{t-1}^{(k)} = 0 \\ r_t^{(F)} & S_{t-1}^{(k)} = -1 \end{cases} \tag{1.2}$$

The corresponding net return of the trading rule k in excess of the benchmark buy-and-hold

[5]This interpretation is also used, for example, in Sullivan, Timmermann, and White (1999) or in Bajgrowicz and Scaillet (2012).

strategy is then given by:

$$r_{EXCESS,t}^{(k)} = r_t^{(k)} - r_t^{(BH)} = \begin{cases} r_t^{(BH)} - r_t^{(F)} & S_{t-1}^{(k)} = 1 \\ 0 & S_{t-1}^{(k)} = 0 \\ r_t^{(F)} - r_t^{(BH)} & S_{t-1}^{(k)} = -1 \end{cases} \tag{1.3}$$

The first performance measure of the trading rule $k \in \{1, \ldots, K\}$ used in this study is then simply the particular trading rule's expected net return (again, in excess of the buy-and-hold strategy):

$$f_M^{(k)} = E\left[r_{EXCESS,t}^{(k)}\right] = E\left[r_t^{(k)}\right] - E\left[r_t^{(BH)}\right] \tag{1.4}$$

Note that this test statistic is constructed in such a way that under market efficiency, each trading rule should not consistently outperform the buy-and-hold strategy, that is, it should hold that $f_M^{(k)} \leq 0$. The natural estimator of this measure, based on the actual observed sequence of signals $\{S_t^{(k)}\}_{t=1}^T$ and returns $\{r_t^{(BH)}\}_{t=1}^T, \{r_t^{(F)}\}_{t=1}^T$ is:

$$\widehat{f}_M^{(k)} = \frac{1}{T}\sum_{t=1}^{T} r_{EXCESS,t}^{(k)}. \tag{1.5}$$

Alternatively, another performance measure that we report in this study is based on the Sharpe Ratio (excess return per unit of risk). Both the benchmark buy-and-hold strategy (represented by the series of returns $r_t^{(BH)}$) as well as the strategy followed by trading rule k (represented by the series of returns $r_t^{(k)}$) may be characterized by the mean return generated by such strategy divided by the standard deviation of these returns. This calculation yields a risk-adjusted performance measure for both strategies, known as the Sharpe Ratio. This measure controls for the possibility that greater expected return is just a compensation for the greater risk of a TTR.[6] We follow standard practice in the literature and calculate the returns in excess of the risk-free rate and obtain the Sharpe Ratios of the trading rule k as well as of the buy-and-hold strategy as follows:

$$\begin{aligned} SR^{(k)} &= \frac{E\left[r_t^{(k)} - r_t^{(F)}\right]}{\sqrt{E\left[\left(r_t^{(k)} - r_t^{(F)}\right)^2\right] - \left(E\left[r_t^{(k)} - r_t^{(F)}\right]\right)^2}} \\ SR^{(BH)} &= \frac{E\left[r_t^{(BH)} - r_t^{(F)}\right]}{\sqrt{E\left[\left(r_t^{(BH)} - r_t^{(F)}\right)^2\right] - \left(E\left[r_t^{(BH)} - r_t^{(F)}\right]\right)^2}}. \end{aligned} \tag{1.6}$$

[6] As noted by Park et al. (2014).

These Sharpe Ratios are naturally estimated by:

$$\begin{aligned}\widehat{SR}^{(k)} &= \frac{\frac{1}{T}\sum_{t=1}^{T}\left(r_t^{(k)} - r_t^{(F)}\right)}{\sigma\left(r_t^{(k)} - r_t^{(F)}\right)} \\ \widehat{SR}^{(BH)} &= \frac{\frac{1}{T}\sum_{t=1}^{T}\left(r_t^{(BH)} - r_t^{(F)}\right)}{\sigma\left(r_t^{(BH)} - r_t^{(F)}\right)},\end{aligned} \tag{1.7}$$

where

$$\sigma\left(x_t\right) = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(x_t - \frac{1}{T}\sum_{j=1}^{T}x_j\right)^2} \tag{1.8}$$

is the standard deviation estimate of the summand in the numerator, which in our case is a time series of strategy returns in excess of the risk-free rate.

Under market efficiency, the risk-adjusted performance of the buy-and-hold strategy should not be smaller than that of any trading rule based strategy. The second performance measure may thus be defined as the difference of the corresponding two Sharpe Ratios. Denoting this measure as $f_{SR}^{(k)} = SR^{(k)} - SR^{(BH)}$, its estimator is given by:

$$\widehat{f}_{SR}^{(k)} = \widehat{SR}^{(k)} - \widehat{SR}^{(BH)} \tag{1.9}$$

We again note that under market efficiency, the following should be satisfied: $f_{SR}^{(k)} \leq 0$.

1.3.3 A Note on Our Universe of Trading Rules

There are hundreds of studies employing various classes of TTRs and there are even more TTRs available from a wider set of sources (such as technical analysis related textbooks) to choose from. Indeed, there is basically an infinite amount of possible rules one could devise and apply to data. In this study, however, we do not follow such aproach. Instead, we focus on a set of relatively simple, well-established and widely studied TTRs described below.

The three classes of trading rules (Moving Average, Trading Range Breakout, Filter) belong to the most widely known and studied in both the academia as well as the commercial investment sector. From the perspective of our research, Table 1.1 provides a representative[7] overview of studies devoted to the analysis of technical trading together with the particular classes of trading rules employed.

It needs to be emphasized that the objective of this study is not to artificially create and discover a new, original technical trading rule that would potentially be successful in terms of

[7]To the best of our knowledge and ability to provide a representative selection of studies both in terms of long-run impact and recency.

Table 1.1: Overview of the literature on profitability of simple technical trading rules.

A checkmark represents the particular TTR being employed in the particular study. Trading rules legend: MA - Moving Average, TRB - Trading Range Breakout, BB - Bollinger bands, F - Filter.

Study	MA	TRB	F	BB	Other
Brock, Lakonishok, and LeBaron (1992)	✓	✓			
Levich and Thomas III (1993)	✓		✓		
Hudson, Dempsey, and Keasey (1996)	✓	✓			
Sullivan, Timmermann, and White (1999)	✓	✓	✓		✓
Leung and Chong (2003)				✓	
Marshall, Cahan, and Cahan (2008)					✓
Lento, Gradojevic, and Wright (2007)				✓	
Neely, Weller, and Ulrich (2009)	✓	✓	✓		
Bajgrowicz and Scaillet (2012)	✓	✓	✓		✓
Yu et al. (2013)	✓	✓			
Park et al. (2014)	✓	✓	✓		✓
Katusiime, Shamsuddin, and Agbola (2015)	✓	✓	✓		
Frömmel and Lampaert (2016)	✓				✓

outperforming the market by possessing a predictable power over the price process. Instead, in order to compare the resilience of multiple stock markets, sampled at several frequencies and analyzed over multiple time periods, we require a set of rules that are reasonably simple and widely applied. The presented evidence highlights the fact that the three classes of trading rules analyzed in this study do conform to such criteria.

1.3.4 Technical Trading Rules

We now introduce the three classes of TTRs (Variable Moving Average, Trading Range Breakout, and Filter) that we employ in this study. We note again that the ultimate goal of all TTRs is to uncover changes in trend of the underlying price discovery process. The following paragraphs describe the way in which this goal is achieved in each class separately.

1. **Variable Moving Average (VMA):** The trend is modelled by comparing two moving averages of past prices, one based on a longer period of time (long moving average, based on the last $N_L \in \mathbb{N}$ price observations) and the other one on a shorter period of time (short moving average, based on the last $N_S < N_L$ price observations). If the former is smaller than the latter, the more recent prices are on average greater than the older prices, which suggests that an upward trend has commenced and one should enter a long position. The logic for downward trend is analogical. We also consider a sensitivity parameter $B > 0$,

which states the minimum possible percentage by which the two moving averages must differ in order for the signals to be triggered. This band of inactivity allows for a certain tolerance level in the difference of the two moving averages, as no signal is generated unless this difference is significant enough so as to exceed the band B and correspondingly, a neutral position is maintained. To summarize, the VMA rule generates a buy (sell) signal if the short moving average rises above (falls below) the long moving average by more than $B \cdot 100\%$.

Each of our VMA rules may be uniquely defined by the three parameters described above. The class of VMA trading rules is therefore parametrized as:

$$VMA(N_S, N_L, B).$$

2. **Trading Range Breakout (TRB):** This rule is based on identification of the so-called trading range, composed of the upper and the lower bounds. We compute these bounds as the maximum and minimum of the past $N \in \mathbb{N}$ price observations. While moving within the trading range, no information about the change in trend is triggered. As soon as the price increases above (declines below) the upper (lower) bound, possibly also adjusted by the band of inactivity $B > 0$, an upward (downward) trend is considered to have commenced and the buy (sell) signal is triggered. The position is maintained until the opposite trend is identified. The two levels that compose the trading range are also called the Resistance and the Support, respectively. Therefore, this class of TTRs is also widely known as the Support-Resistance class.

   Uniquely defined by the two parameters described above, we denote the class of TRB rules as

$$TRB(N, B).$$

3. **Filter (F):** This class of trading rules focuses on the magnitude of price changes. In particular, the two percentage filters are used to determine the values of price changes that are considered to be significant in terms of triggering a signal for opening and closing a position. These changes are always evaluated with respect to a reference price. If the price rises above (falls below) the reference price by more than the value of the opening filter ($x > 0$), a long (short) position is entered and the reference price is defined as the most recent maximum (minimum) price. If the price then falls below (rises above) this reference price by more than the value of the closing filter (y), a position is closed and a neutral position is taken and the new reference price is defined as the price prevailing at the time of entering this neutral position. The Filter rule is similar to the previous rules in the sense that it aims to uncover the underlying bullish and bearish trends in the price

process and trade on this information accordingly, entering a long position in an upward trend and a short position in a downward trend, while ignoring small price fluctuations by using appropriate values of the two filters. If these values successfully correspond to the timing of the market bullish/bearish behaviour, a profitable strategy is obtained. From the behavioural perspective, the two filters may be used to emulate the tolerance levels of market participants in the process of forming their expectations about the nature of the market movement in the future.

Uniquely defined by the two parameters described above, we denote the class of Filter rules as

$$F(x, y).$$

The particular values of the parameters used, as well as the exact algorithmic prescriptions of each class of TTRs are provided in Appendix 1.A. Overall, we employ 296 VMA rules, 176 TRB rules and 144 F rules. Therefore, the total number of individual TTRs that the "universe of trading rules" considered in our study contains is 616.

For an illustration, we demonstrate the underlying mechanics for each class of TTRs on Figure 1.1. We have simulated a Geometric Brownian Motion (GBM) path with trend 0 and volatility 0.01 consisting of 500 observations and applied three particular (randomly chosen out of our universe) TTRs: $VMA(5, 30, 0.005)$, $TRB(40, 0.01)$ and $F(0.03, 0.02)$. The long, short and neutral positions generated by these rules are represented with the blue, red and white background, respectively.

Additionally, we repeat the entire procedure for 1000 random trajectories of GBM. The histograms of the two performance measures, namely, the excess mean returns and excess Sharpe Ratios for the $F(0.03, 0.02)$ rule are displayed on Figure 1.2. The thick red lines correspond to the particular GBM path displayed on Figure 1.1.

We may observe that the distributions are centered around (or mildly below) zero, suggesting that on average, the three TTRs are not informative about the changes in trend. This is not surprising, since the simulated prices are martingales. However, for our particular realization on Figure 1.1, the Filter rule yields purely by chance considerably positive values of both performance measures. It is these random artifacts that the data-snooping tests are meant to uncover.

1.3.5 Testing for Data-Snooping

If the two performance measures described above (the mean return and the Sharpe Ratio) have positive values, we have an indication that the trading rule k outperforms the benchmark buy-and-hold strategy either from the mean return or from the risk-adjusted Sharpe Ratio

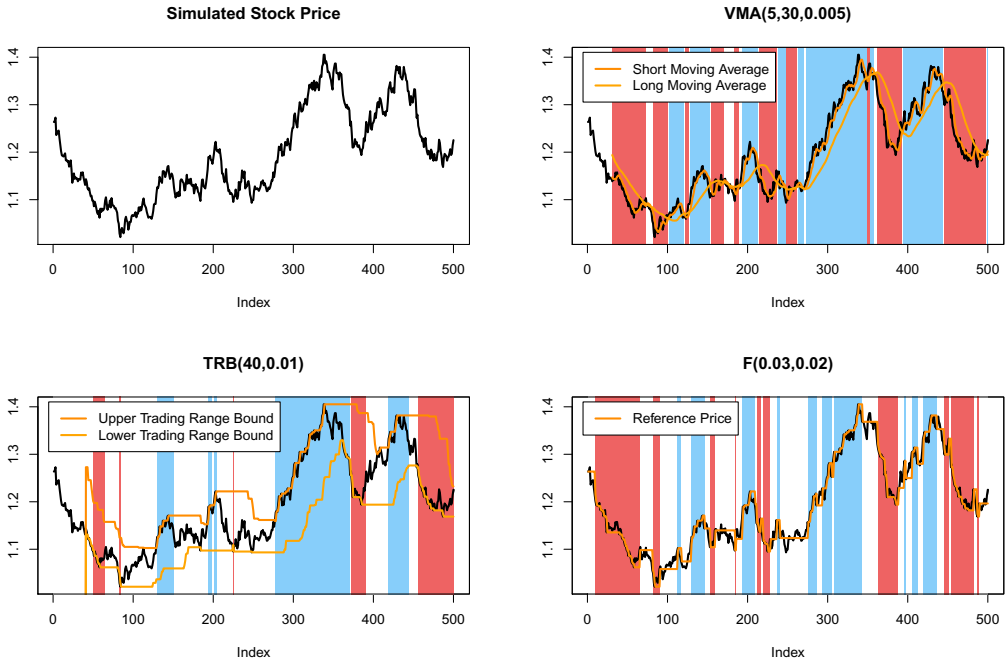


Figure 1.1: The four figures display the simulated stock price of length 500 (upper left figure) and the functionality of the three class of TTRs: Variable Moving Average (upper right figure), Trading Range Breakout (lower left figure) and Filter (lower right figure). White, blue and red background represent trading rule signals at any point in time: neutral, buy and sell, respectively.

perspective (or both). However, an evaluation of the statistical significance of such result is important. If one tries sufficient number of different trading rules, some of them will indeed appear as outperforming, even though such result would simply be a statistical artifact, while the trading rule would not be bearing any true predictive power over the benchmark strategy.

The dangers of data-snooping (as such phenomenon of using the same dataset in the process of model selection multiple times is named) have been acknowledged in the literature. Brock, Lakonishok, and LeBaron (1992) evaluate the obtained results by fitting several null models to the returns data and performing a bootstrap resampling of the residuals. However, as they acknowledge, they are not able "to compute a comprehensive test across all rules. Such a test would have to take into account the dependencies between results for different rules." (Brock, Lakonishok, and LeBaron (1992), p. 1743). Nevertheless, in each study that estimates the performance of multitude of trading rules, such test is of extreme importance.

The most widely used procedure designed to adress the issue of data-snooping is the White (2000) Reality Check bootstrap-based test and is applied for the first time by Sullivan, Timmermann, and White (1999) and later by numerous researchers, for example, Marshall, Cahan, and Cahan (2008), Bajgrowicz and Scaillet (2012) and Park et al. (2014). It tests the null hypothesis that the best performing trading rule (out of a pre-defined set, or "universe", of trading rules) is no better than a given benchmark model. As Sullivan, Timmermann, and White (1999) state,

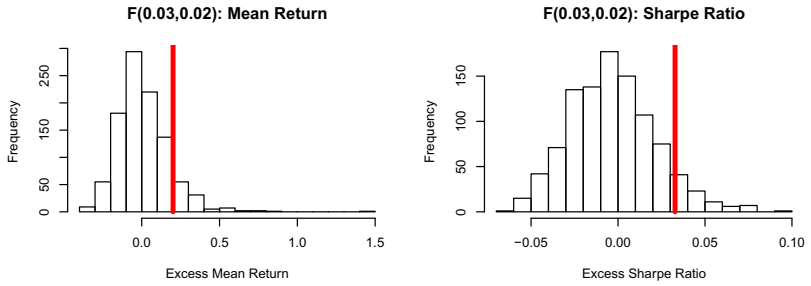


Figure 1.2: Histograms of mean excess returns and Sharpe Ratios for 1000 random stock price trajectories of length 500 for the $F(0.03, 0.02)$ trading rule. The vertical red lines represent the values observed for the particular trajectory displayed on Figure 1.1.

"the idea is to evaluate the distribution of a suitable performance measure giving consideration to the full set of models that led to the best-performing trading rule." (Sullivan, Timmermann, and White (1999), p. 1650).

Considering the mean return performance measure $f_M^{(k)}$ defined above, the null hypothesis of White Reality Check test may be formulated as follows:

$$H_0 : \max_{k=1,\ldots,K} f_M^{(k)} \leq 0 \tag{1.10}$$

If we reject the null hypothesis, we will possess an evidence that the best trading rule in our universe of trading rules (containing K rules) indeed outperforms the benchmark strategy (from a mean return perspective) in a statistically significant manner.

The test statistic is equal to:

$$V = \max_{k=1,\ldots,K} \{\sqrt{T}\hat{f}_M^{(k)}\} \tag{1.11}$$

White (2000) shows that the probability distribution of this test statistic under the null hypothesis may be approximated by performing the stationary bootstrap algorithm, originally introduced by Politis and Romano (1994). The main idea of this algorithm is to re-sample blocks of original time series of returns of varying length. The exact algorithm is described in Appendix 1.A. For each bootstrap simulation b, the corresponding test statistic is calculated as follows:

$$V_b^* = \max_{k=1,\ldots,K} \{\sqrt{T}\left(\hat{f}_{M,b}^{(k)*} - \hat{f}_M^{(k)}\right)\}, \quad b = 1, \ldots, B \tag{1.12}$$

where $\hat{f}_{M,b}^{(k)*}$ is the mean return performance measure estimate corresponding to the k-th trading rule applied to the b-th bootstrapped time series of returns. Repeating the stationary bootstrap B times yields B values of $V_b^*, b = 1, \ldots, B$. The p-value corresponding to the null hypothesis

is then calculated as:

$$\text{p-value} = \frac{\sum_{b=1}^{B} I(V_b^* > V)}{B}, \tag{1.13}$$

where $I(\cdot)$ is the indicator function, equal to 1 if the condition in the parentheses is true and zero otherwise. This procedure allows us to statistically evaluate whether the best performing rule in our universe of trading rules (containing 616 individual rules) possesses predictive power over the benchmark buy-and-hold strategy. Under market efficiency, such predictive power should not be present and the null hypothesis should not be rejected.

The above procedure used estimates of the mean return performance measure, $\hat{f}_M^{(k)}$. Alternatively, and analogically, we also consider the Sharpe Ratio performance measure $\hat{f}_{SR}^{(k)}$.

Despite the wide usage of White's Reality Check procedure in the literature, we note that the method possesses several drawbacks that might to a certain extent affect our results. First, by construction, the Reality Check only tests whether the single best performing trading rule beats the benchmark strategy after accounting for data-snooping. Therefore, if there are other well performing rules in our universe, these are not considered in the procedure. Moreover, as argued by Hansen (2005), the Reality Check may perform poorly in the cases where the universe of trading rules contains many rules that perform poorly and some that perform well.[8] In this case, the configuration of the test is "the least favorable to the alternative hypothesis" (Park et al. (2014)). These facts need to be acknowledged, since we cannot rule out the possibility of such composition of trading rules prevailing also in our universe.

1.4 Data

Potential profitability of technical analysis is investigated for financial markets of the following seven countries: Austria (ATX), the Czech Republic (PX50), Germany (DAX), Hungary (BUX), Poland (WIG), Slovenia (SBITOP), and the United States (SPX).[9] The original data is sampled at the 5-minute frequency, sourced from Bloomberg and covers periods January 2008 - July 2013 for the ATX, SBITOP, and SPX and January 2008 - December 2015 for the BUX, DAX, PX, and WIG indices.

We use the original 5-minute frequency data to create the three frequencies used in our study: 15-minute, hourly, and daily. We thus obtain 21 time series of closing prices. Additionally, since we wish to observe the time development of the results, we divide each time series into non-overlapping subsamples. This division is carried out for each frequency separately as follows:

[8] Hansen (2005) also proposes an alternative data snooping bias test modification, known as the Superior Predictive Ability test. In this study, we apply the original White's Reality Check. As documented in, for example, Park et al. (2014), the two procedures do not report dramatically different results. Similarly, an alternative data snooping testing procedure is proposed by Hsu, Hsu, and Kuan (2010).

[9] Each country's financial market is proxied by the corresponding stock market index, the abbreviation of which is stated in parentheses.

1. **Daily:** for the daily data, we set the length of one subsample to be equivalent to **52 weeks** (approximately one year) worth of data. Consequently, one subsample of daily data consists of approximately 250 observations.

2. **Hourly:** for the hourly data, one subsample is equivalent to **12 weeks** (approximately one quarter-year) worth of data. This yields approximately 420 hourly observations per subsample.

3. **15-minute:** for the data sampled at 15-minute frequency, one subsample is equivalent to **4 weeks** (approximately one month) worth of data, yielding approximately 560 observations per subsample.

We choose different subsample sizes for different frequencies for several reasons: for each subsample, it is desirable to have at least 100 (ideally more) closing price observations to see how the technical trading rules perform. Moreover, we wish to avoid comparing the results of subsamples that have dramatically different sample sizes - therefore, our division into subsamples yields sample sizes that are not orders of magnitude different from each other. Figure 1.3 provides a graphical illustration of the subsamples creation process described in this section. The first price of each subsample is the first price at which we assume an investment position may be taken - if a trading rule requires the knowledge of a certain number of previous prices (as, for example, the variable moving average rule always does), these prices are not included in the subsample, but we naturally use them for calculation of trading signals as necessary.

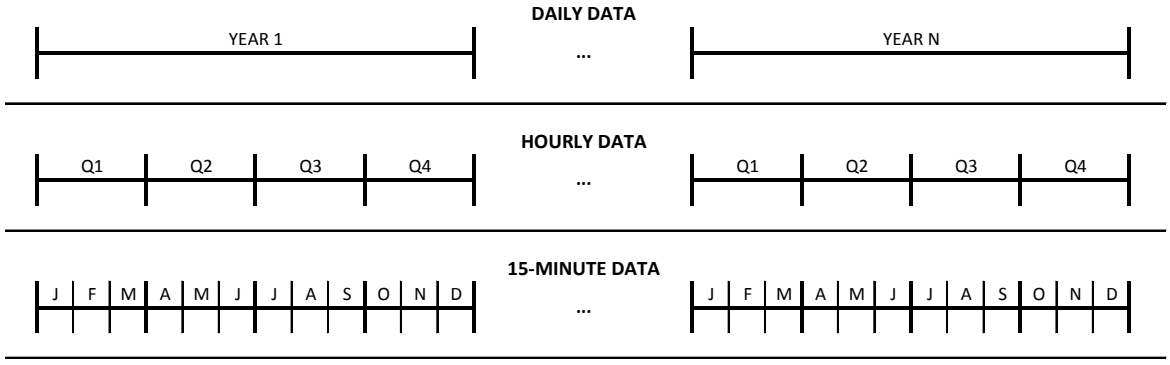


Figure 1.3: Scheme of creating subsamples. Each vertical line represents a division into subsamples. One subsample is equivalent to approximately 1 year for the daily data (top panel), 1 quarter for the hourly data (middle panel) and 1 month for the 15-minute data (bottom panel).

Figure 1.4 displays the price development of the stock market indices used in this study, while Table 1.2 displays the summary statistics of the corresponding daily log-returns. We observe that even though all regions share a strong downward trend during 2008, corresponding to the global financial crisis, there are considerable differences in the post-crisis period price

development. While the S&P 500 as well as the German DAX rebounded fairly quickly, the central European indices exhibit range-like movement, while the Slovenian SBITOP continued the slight downward drift. The Slovenian index also exhibits considerably smaller volatility than all the other indices. The normality of daily log-returns of all indices is rejected by the Jarque-Bera test. The log-returns across all indices and subsamples exhibit excess kurtosis and are not symetrically distributed, as confirmed by the skewness statistic, which suggests, with only a few exceptions, negatively skewed log-return distributions.

In our calculations, we also use the data on the risk-free rates. For that purpose, we use the short term interest rate data available at the OECD website.[10]

We use the data corresponding to the currencies of the countries that are represented by the stock market indices. The data is provided at the monthly frequency. We convert the annualized simple interest rate values $R_t^{(F)}$ to the corresponding one-period logarithmic interest rate $r_t^{(F)}$ using the formula:

$$r_t^{(F)} = \frac{\log(1 + R_t^{(F)})}{M}, \tag{1.14}$$

where $M = 252$ for the daily frequency calculations, $M = 252 \cdot 8$ for the hourly frequency calculations and $M = 252 \cdot 8 \cdot 4$ for the 15-minute frequency calculations.

We also provide additional information (as of June, 2017) about the five Central and Eastern European indices:

- **ATX (Austria):** ISIN: AT0000999982. The index represents the 20 largest blue chip companies traded on Wiener Börse. The index is followed by the corresponding Exchange Traded Fund (ETF) iShares ATX UCITS (ISIN: DE000A0D8Q23) that is traded at Vienna and Frankfurt exchanges. The largest constituent of the index is Erste Bank with 19.55% weight. Index futures and options are traded at the Frankfurt Exchange (Eurex).

- **BUX (Hungary):** ISIN: XC0009655090. The index representing the 25 largest blue chip Hungarian companies traded on Budapest Stock Exchange (BSE). The largest constituent is OTP Bank (33.03%). It is a tradeable index, its derivative contracts are frequently traded at BSE.

- **PX (Czech Republic):** ISIN: XC0009698371. Opened on April 5, 1994, based on the largest companies traded on Prague Stock Exchange. The largest constituent is Erste Group Bank (21.67%). The index has its own corresponding ETF (managed by Horizon, ISIN: BE6224091866), though, to the best of our knowledge, this ETF is a passively managed fund.

[10] `https://data.oecd.org/interest/short-term-interest-rates.htm`

- **SBITOP (Slovenia):** ISIN: SI0026109882. Launched on March 31, 2006, the index consists of 11 largest traded companies on Ljubljana Stock Exchange. The largest constituent is KRKA (30%). The index is designed as a tradeable index and is available to financial institutions as an underlying in derivatives products. To the best of our effort, we have not found an ETF that would directly follow this index.

- **WIG (Poland):** ISIN: PL9999999987. Also denoted as WIG20, the Polish index was introduced on April 16, 1991. The largest constituent is PKN Orlen (9.99%). Index futures (since 1998) and options (since 2003) are traded on Warsaw Stock Exchange. The corresponding index ETF is traded as well as the LYXOR WIG20 ETF (ISIN: LU0459113907).

We note that Vienna Stock Exchange also acts as exclusive sales agent for the indices of Prague and Ljubljana Stock Exchange. Moreover, trading at stock exchanges in Vienna (since 1999), Prague (since November 2012), Budapest (since December 2013) and Ljubljana (since December 2010) is organized via the Xetra electronic trading system.

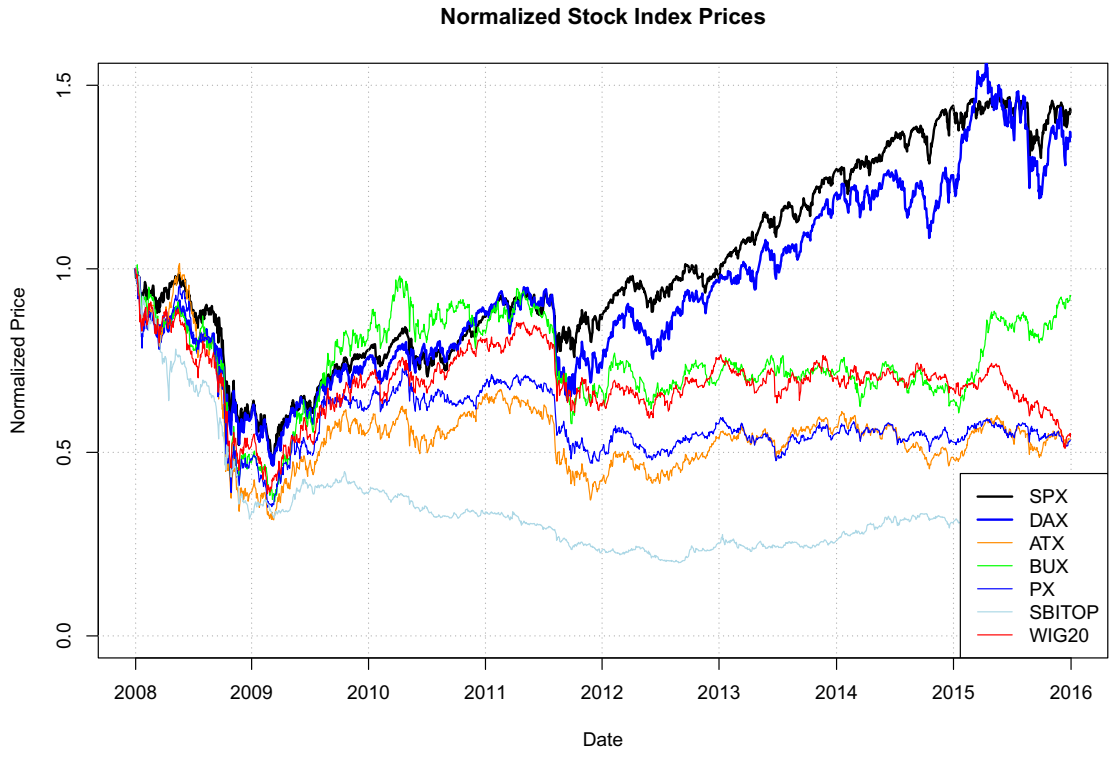


Figure 1.4: Daily normalized closing prices of the stock market indices considered in this study. Each original time series is normalized by dividing each element by the first available corresponding closing price (as of January 2, 2008).

Table 1.2: Summary statistics of daily log-returns.

The table contains several summary statistics of the stock market closing prices used in this study for seven different market indices. We report summary statistics of the daily log-returns for the entire available sample as well as for three subsample periods. $Q(1)$ denotes the first-order autocorrelation coefficient. JB denotes the p-value of the Jarque-Bera normality test.

	ATX	BUX	DAX	PX	SBITOP	SPX	WIG
Region	Austria	Hungary	Germany	Czechia	Slovenia	USA	Poland
Start	2008/01/02	2008/01/02	2008/01/02	2008/01/02	2008/01/03	2008/01/02	2008/01/02
End	2013/07/19	2015/12/04	2015/12/04	2015/12/04	2013/07/05	2013/07/19	2015/12/04
# daily	1378	1980	2016	1990	1374	1396	1986
# hourly	12393	17073	18126	16426	6510	11139	17108
# 15-min	49572	68292	72504	65704	26040	44556	68432
Daily log-returns summary statistics:							
Entire sample							
Mean	-0.0005	0.0000	0.0001	-0.0003	-0.0010	0.0001	-0.0003
Stdev	0.0199	0.0173	0.0154	0.0160	0.0135	0.0163	0.0154
Skewness	-0.1040	-0.0467	0.0735	-0.5162	-0.3954	-0.2767	-0.2984
Kurtosis	6.5934	10.1825	8.4119	17.8375	9.4200	10.8346	6.7564
Q(1)	0.0796	0.0524	0.0086	0.0594	0.1302	-0.1279	0.0443
JB	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Jan 1, 2008 - Dec 31, 2009							
Mean	-0.0012	-0.0004	-0.0007	-0.0010	-0.0019	-0.0005	-0.0007
Stdev	0.0262	0.0255	0.0214	0.0255	0.0177	0.0222	0.0230
Skewness	-0.0477	-0.0550	0.3334	-0.3741	-0.2820	-0.1033	-0.1249
Kurtosis	5.0093	6.8756	7.2979	10.3163	7.4945	7.1016	4.0982
Q(1)	0.0618	0.1038	-0.0389	0.0812	0.1896	-0.1335	0.0568
JB	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Jan 1, 2010 - Dec 31, 2011							
Mean	-0.0006	-0.0004	0.0000	-0.0005	-0.0010	0.0002	-0.0003
Stdev	0.0169	0.0172	0.0150	0.0138	0.0092	0.0141	0.0142
Skewness	-0.0237	0.1527	-0.1826	-0.3157	-0.5359	-0.4792	-0.4749
Kurtosis	5.5146	7.1066	5.1916	6.3805	7.8877	9.8116	6.1868
Q(1)	0.1150	-0.0328	0.1081	0.0492	0.0258	-0.1437	0.0222
JB	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Jan 1, 2012 - End of sample							
Mean	0.0004	0.0003	0.0006	0.0001	0.0001	0.0007	-0.0002
Stdev	0.0130	0.0113	0.0117	0.0095	0.0117	0.0078	0.0104
Skewness	-0.1148	-0.0381	-0.2026	-0.1757	-0.1740	-0.1873	-0.4473
Kurtosis	3.9376	4.8798	4.2841	4.5660	4.3071	4.0028	5.4965
Q(1)	0.0762	0.0237	-0.0068	-0.0137	0.0223	-0.0181	0.0335
JB	0.0006	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000

1.5 Results

1.5.1 Performance and the Reality Check Evaluation

We apply the 616 technical trading rules contained in our universe to each of the seven stock market indices: Austrian (ATX), Hungarian (BUX), German (DAX), Czech (PX), Slovenian (SBITOP), American (SPX) and Polish (WIG). For each index, we consider three different frequencies of closing prices, approximating the trading strategies of three different types of investors, based on whether they evaluate (and potentially change) their positions on a daily, hourly, or 15-minute basis. For each index and each frequency, we apply the TTRs to every subsample created as described earlier.

For each index, for each frequency, for each subsample, and for each trading rule, we evaluate the performance of the trading rule using the two measures introduced earlier: the mean return and the Sharpe Ratio. Both measures are evaluated against the benchmark buy-and-hold strategy, consisting of entering a long position in the index at the beginning of a subsample and holding that position until the end of the subsample. If one of these two measures is greater for a particular trading rule than it is for the benchmark strategy, we have an indication that this trading rule outperforms the market.

We also take into account the possibility of data-snooping bias and carry out White's Reality Check procedure to evaluate the statistical significance of our results. For each index, for each frequency, and for each subsample, White's Reality Check is performed by testing the null hypothesis that the best performing trading rule in our universe is not better than the benchmark. We perform this Reality Check twice: both for the mean return and for the Sharpe Ratio entering as the performance measure used to calculate the test statistics.

Benchmark S&P 500 Index

Because of the large quantitative nature of our study, we report the results in graphical form. Figure 1.5 shows the results for one stock market index and the first performance measure: the S&P 500 and the mean return performance measure. The S&P 500 is the most widely studied and developed stock market of the United States, and thus serves as our benchmark index. We describe how these graphical results are meant to be read in the following paragraphs.

There are 6 plots displayed in Figure 1.5. The first plot shows the performance of each of the 616 TTRs applied to the closing prices sampled at daily frequency. The horizontal axis represents the beginning of each subsample, while the vertical axis represents the terminal value obtained by investing $1 in the beginning of the subsample period and following the corresponding trading strategy. For each subsample period (the beginning of which is labeled on the x-axis), we have 617 different values of these terminal wealths - one for each of the 616 TTRs and one for the

benchmark buy-and-hold strategy (shown as the thicker black line).

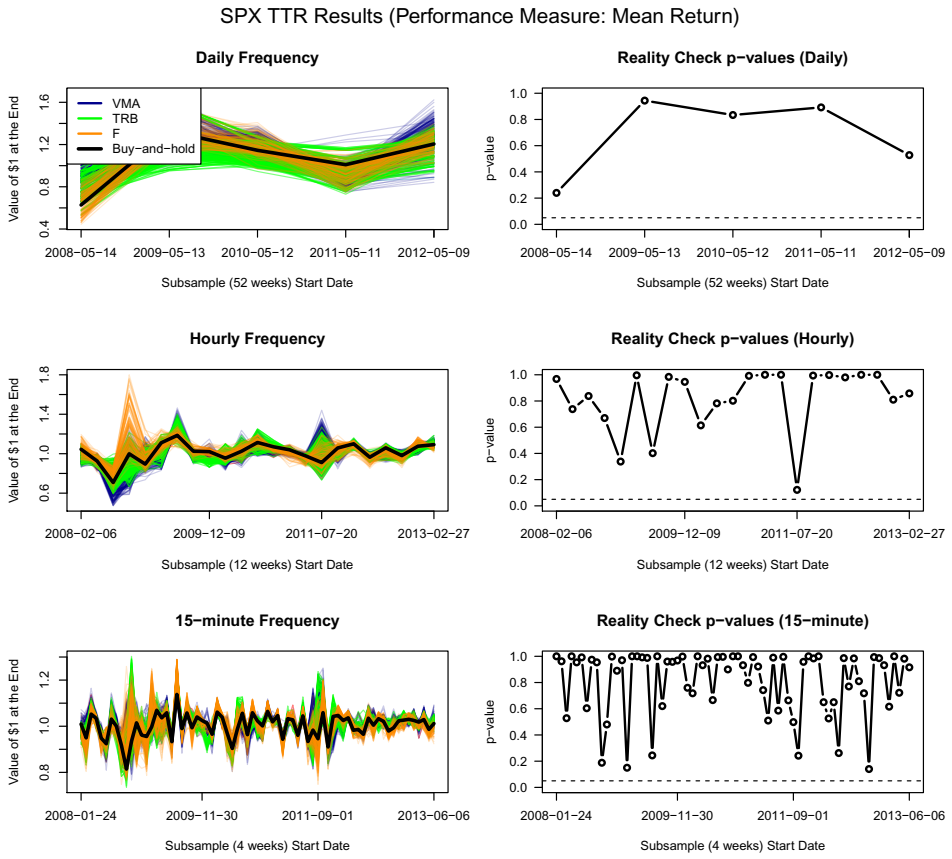


Figure 1.5: Technical Trading Rules evaluation results summary for the S&P 500 index (SPX) and the mean return performance measure.

The TTRs are represented in three different colors, based on the particular class of trading rules (dark blue for the Variable Moving Average, green for the Trading Range Breakout, and orange for the Filter rule). This terminal wealth evaluation is equivalent to the mean return performance measure (there is one-to-one mapping between the terminal value of $1 invested and the mean return).

The right-hand side plots in Figure 1.5 display the p-values of White's Reality Check procedure applied to each subsample at different frequencies. The dashed line represents the 5 percent level of significance. Any p-value below this level corresponds to a subsample during which the best performing TTR contained statistically significant (at the 5 percent level) predictive power over the price process.

The two plots in the first row show the results for daily frequency data. The plots in the second and the third rows follow the same logic, but display the results for hourly and 15-minute frequency data, respectively.[11]

[11]We remind the reader that each daily frequency subsample contains 52 weeks (1 year) of data, while

The results for the S&P 500 index are largely unsurprising. For both daily and hourly frequencies, the vast majority of TTRs underperformed the benchmark buy-and-hold strategy. Despite the index exhibiting pronounced trends during our sample (corresponding to the post-2008 bearish market, followed by gradual post-2009 recovery), few trading rules managed to outperform the benchmark strategy for each subsample period. Moreover, the daily and hourly White's Reality Check test p-values are all well above the 0.05 level for all subsamples.

The results are qualitatively only slightly different for the 15-minute frequency data. Here, in several subsamples, a large fraction of the TTRs that were outperforming the index appear. Moreover, the White's Reality Check p-values are smaller in general, though still mostly far from the critical 5-percent level.

Overall, the United States stock market, as proxied by the S&P 500 index, exhibits results that are consistent with the EMH, since we fail to reject the null hypothesis of no excess performance of technical analysis in any case. Though some of our TTRs manage to outperform the market, we conjecture that such observations are due to chance, and no TTR bears true predictive power over the price process. We therefore turn now to the European stock market indices.

We note that the mean return and the Sharpe Ratio results are qualitatively very similar, not only for the US index, but for all markets and frequencies considered in this study. Therefore, for the sake of brevity, we report the rest of the results only for the mean return performance measure. The risk-adjusted Sharpe Ratio results are reported in Appendix 1.A for all indices.

European Indices

Overall, the results for the European stock indices are not exceptionally distant from those of our benchmark S&P 500 index, especially regarding the daily and hourly frequency data. It is possible, however, to observe certain systematic differences between the American and German indices and the remaining five European indices. These differences especially manifest in the data sampled at the 15-minute frequency.

Considering the indices individually, we observe that the ATX index (Figure 1.6) was frequently beaten by the VMA trading rules (blue lines), while the TRB and F rules performed comparably to the buy-and-hold strategy. However, in all of the daily and hourly subsamples, we fail to reject the hypothesis that the performance of the best trading rule is not due to the data-snooping phenomenon, although in the first daily subsample and two of the hourly subsamples, the p-values are very close to the critical level. Looking at the hourly frequency results, the p-values are, in general, considerably smaller than for the S&P, though they still fail to reject

each hourly frequency subsample contains 12 weeks (one quarter) of data, and each 15-minute frequency subsample contains 4 weeks (one month) of data. For each frequency, the adjacent subsamples are non-overlapping.

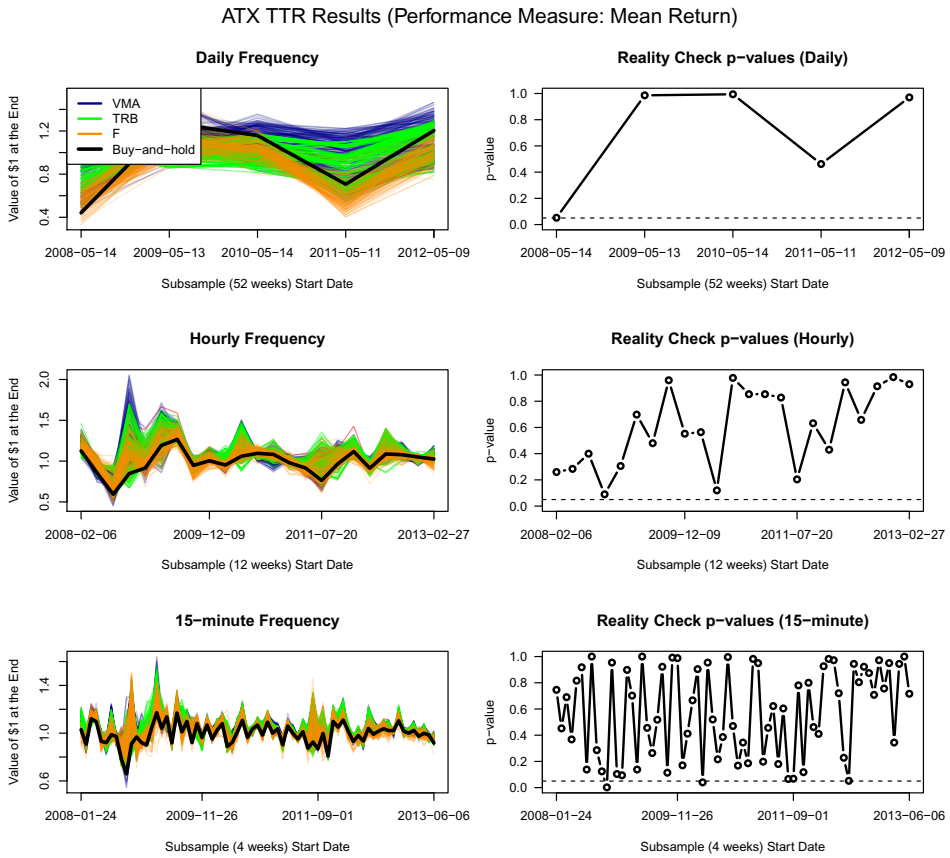


Figure 1.6: Technical Trading Rules evaluation results summary for the Austrian index (ATX) and the mean return performance measure.

the null hypothesis.

At the 15-minute frequency, however, there are three subsamples that reject the null hypothesis of no abnormal performance of the best trading rule.

The results for the Hungarian BUX market (Figure 1.7) are qualitatively very similar to the ATX index. After accounting for data-snooping, our trading rules are not able to outperform the index at the daily and hourly frequencies, though there are three significant subsamples at the 15-minute frequency.

The German DAX index (Figure 1.8), which is arguably the most liquid and developed stock market in continental Europe, provides evidence consistent with the EMH. Compared to the other European indices, we observe much higher occurrence of p-values very close to 1, and notably smaller number of (nearly) significant White's Reality Check results. We fail to reject the null hypothesis in each subsample and at all three frequencies.

The results of the Czech PX index (Figure 1.9) suggest that there existed periods of market inefficiency during which the stock market could be exploited by using technical analysis. PX is the only stock market index, for which there exists a daily frequency subsample (starting on May 14, 2008) that rejects the null hypothesis of White's test. We observe significant Reality Check

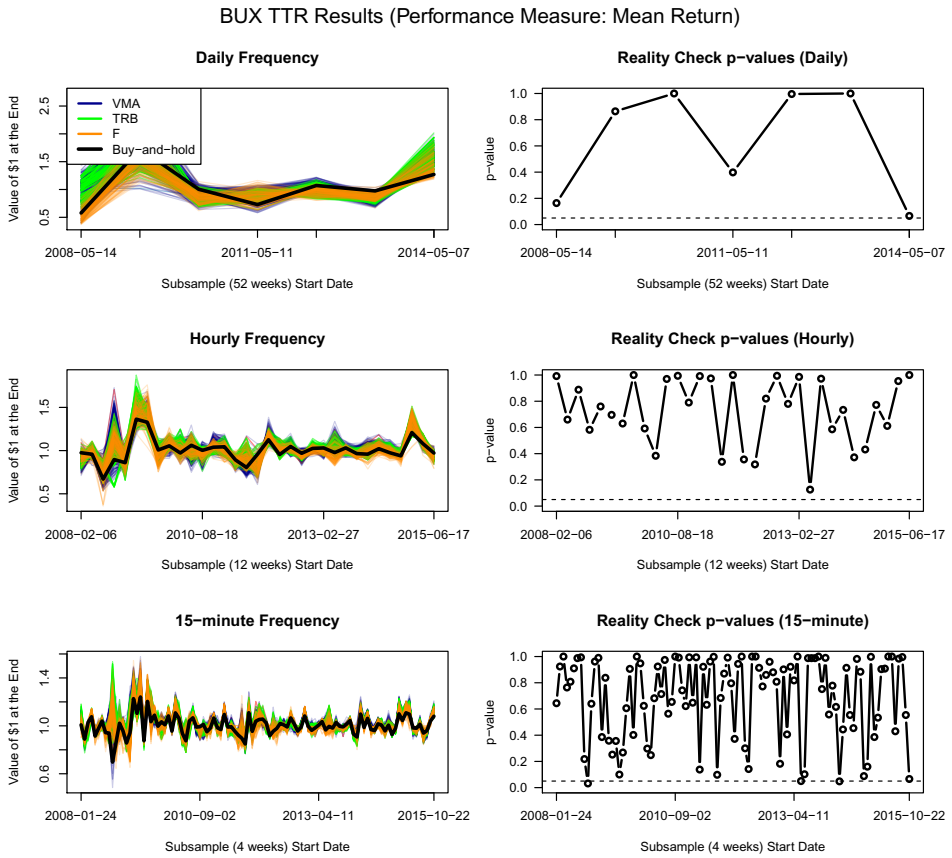


Figure 1.7: Technical Trading Rules evaluation results summary for the Hungarian index (BUX) and the mean return performance measure.

p-values at all three frequencies. However, the number of significant p-values still represents only a small fraction of the subsamples considered.

The Slovenian SBITOP (Figure 1.10) contains the largest number of subsamples in which we reject the Reality Check null hypothesis. Even though the market appears to be efficient at daily frequency, we reject the null hypothesis for three hourly frequency subsamples and for eight 15-minute frequency subsamples. These subsamples are present mostly during the first half of the period observed, in particular, during the last quarter of 2008, the period corresponding to the most intense financial crisis stress.

Finally, the Polish WIG index (Figure 1.11) again provides evidence mostly in favor of the Efficient Markets Hypothesis. We observe a high resemblance between the WIG and DAX indices. With the exception of one instance in the 15-minute frequency data (again corresponding to autumn of 2008), all Reality Check p-values are above 0.05 and overall maintain higher values, especially for daily and hourly frequencies.

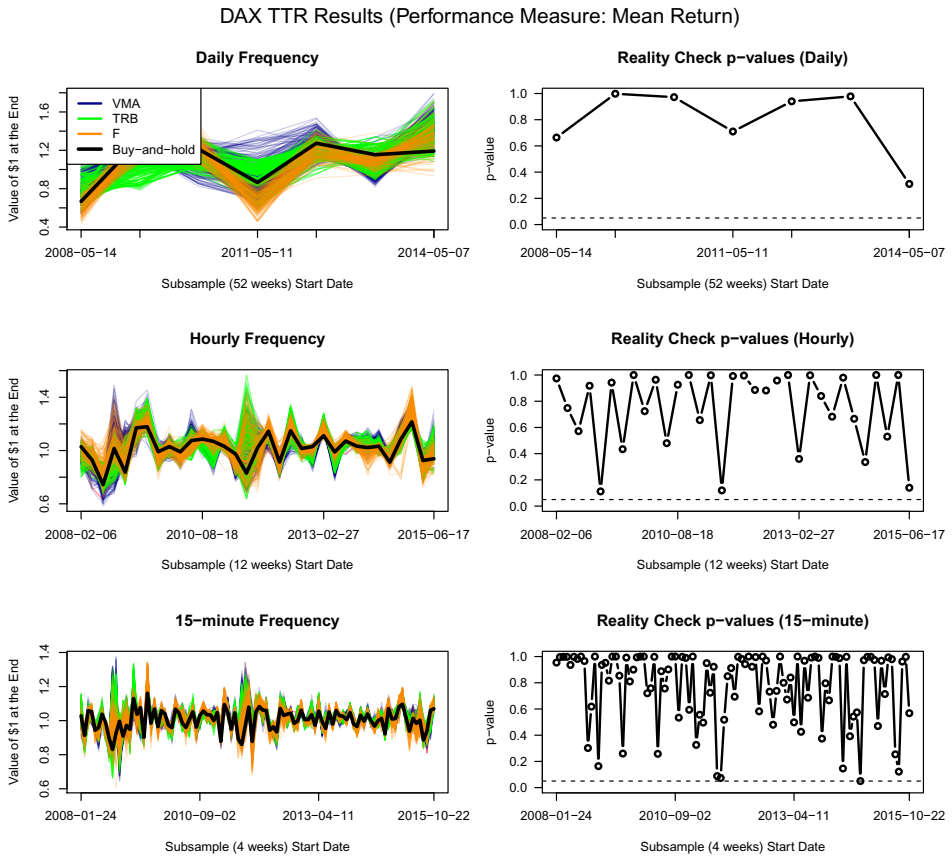


Figure 1.8: Technical Trading Rules evaluation results summary for the German index (DAX) and the mean return performance measure.

Table 1.3: Number of statistically significant Reality Check subsamples.

The table contains the number of subsamples, for which the Reality Check procedure applied to the mean return performance measure rejects the null hypothesis of no abnormal performance of the best trading rule.

Frequency	SPX	DAX	ATX	BUX	PX	SBITOP	WIG
Daily	0	0	0	0	1	0	0
Hourly	0	0	0	0	2	3	0
15-minute	0	0	3	3	1	8	1

Summarizing the results presented in this section (see also Table 1.3), we find evidence that all the financial markets (as proxied by their corresponding market indices) have been efficient with respect to the set of relatively simple trading rules providing signals using daily frequency data during the period covered in our sample. Moreover, the stock market indices of the United States and Germany (the two most developed markets in our sample) also exhibit the same results when both types of higher frequency data (hourly and 15-minute) are considered. For the remaining five CEE markets, we find some evidence of the existence of trading rules that possess

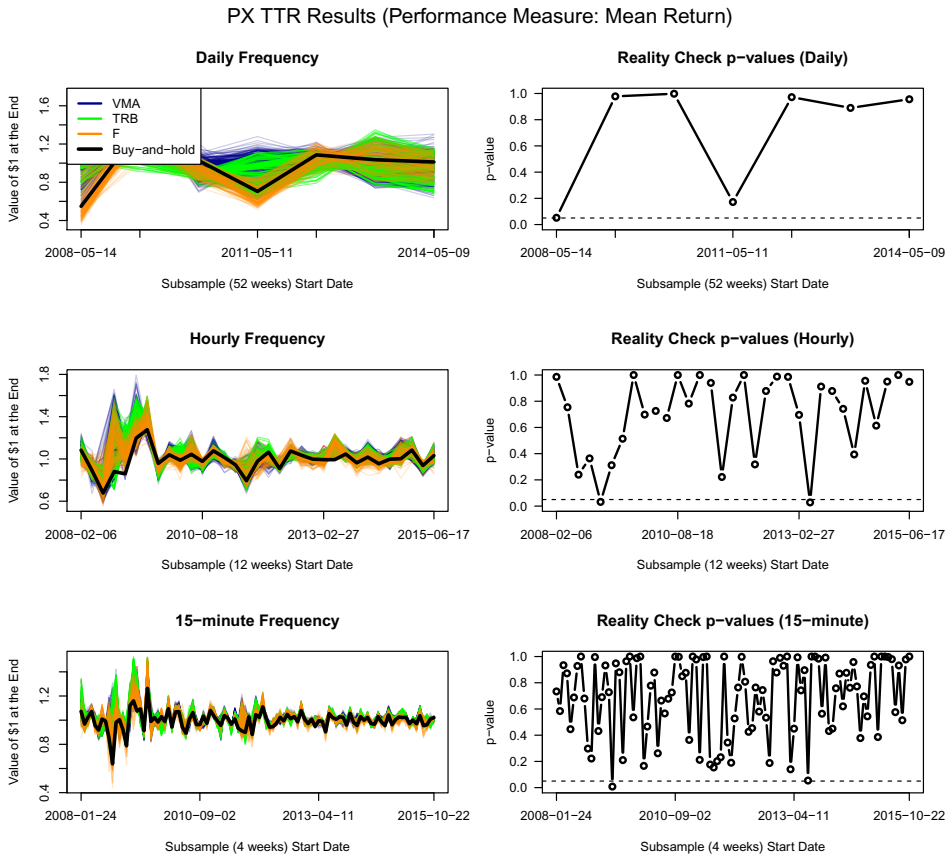


Figure 1.9: Technical Trading Rules evaluation results summary for the Czech index (PX) and the mean return performance measure.

true predictive power over the indices for several hourly and 15-minute subsamples. However, these subsamples do not occur frequently and are always followed by periods during which evidence of market efficiency (with respect to our trading rules) again prevails. Nevertheless, for these five stock markets, we observe periods characterized by mixed evidence of market efficiency.

We must note, however, that to claim that a certain market is truly inefficient, one needs to be able to actively trade on this market, while simultaneously bearing transaction costs that do not wipe out all the excess return yielded by a technical trading strategy. Considering the Slovenian index as an example, even though certain exchange traded funds (ETFs) exist with Slovenian exposure (such as the iShares MSCI Frontier 100, which is exposed to the emerging markets), to the best of our knowledge, no ETF exists that would directly be following the SBITOP index (see section 1.4 for a discussion). It is therefore possible that trading on this index would be accompanied with considerably high transaction costs that could potentially wipe out all profits included in a trading strategy, as an investor would have to rebalance every asset included in this stock index individually, especially considering that such actions would potentially need to be carried out multiple times per day.

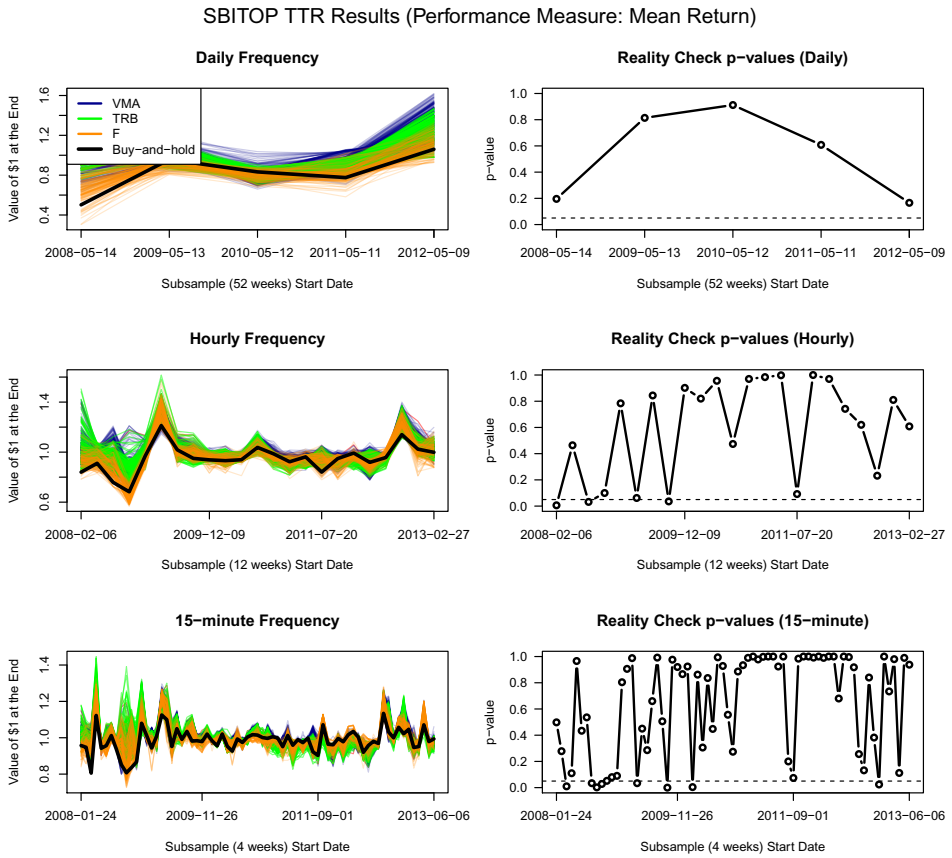


Figure 1.10: Technical Trading Rules evaluation results summary for the Slovenian index (SBITOP) and the mean return performance measure.

1.5.2 A Thought Experiment: the "Best Rule" Investment Strategy and Transaction Costs

The non-overlapping subsamples structure of our analysis enables us to observe how all the results evolve over time. One way in which it is possible to exploit such results is to evaluate how a hypothetical investor following a certain set of rules would perform over time. Observing the left-hand side graphs shown on Figures 1.5 - 1.11, we note that, for all subsamples (disregarding frequency), a multitude of TTRs do outperform the corresponding benchmark buy-and-hold strategy. Despite the fact that White's Reality Check procedure does not frequently qualify a performance of the best trading rule as immune to data-snooping, we note that it is not unusual to observe two (or more) adjacent subsamples during which the same trading rule outperformed the market. Visually, this may be expressed as observing a fixed colored line above the thick black line for a multitude of adjacent subsamples. We are therefore interested in analyzing how a hypothetical investor observing outperforming rules *in the near past*, who follows them *in the near future* would perform.

We devise the following, simplest possible strategy (and we refer to it as the "Best Rule"

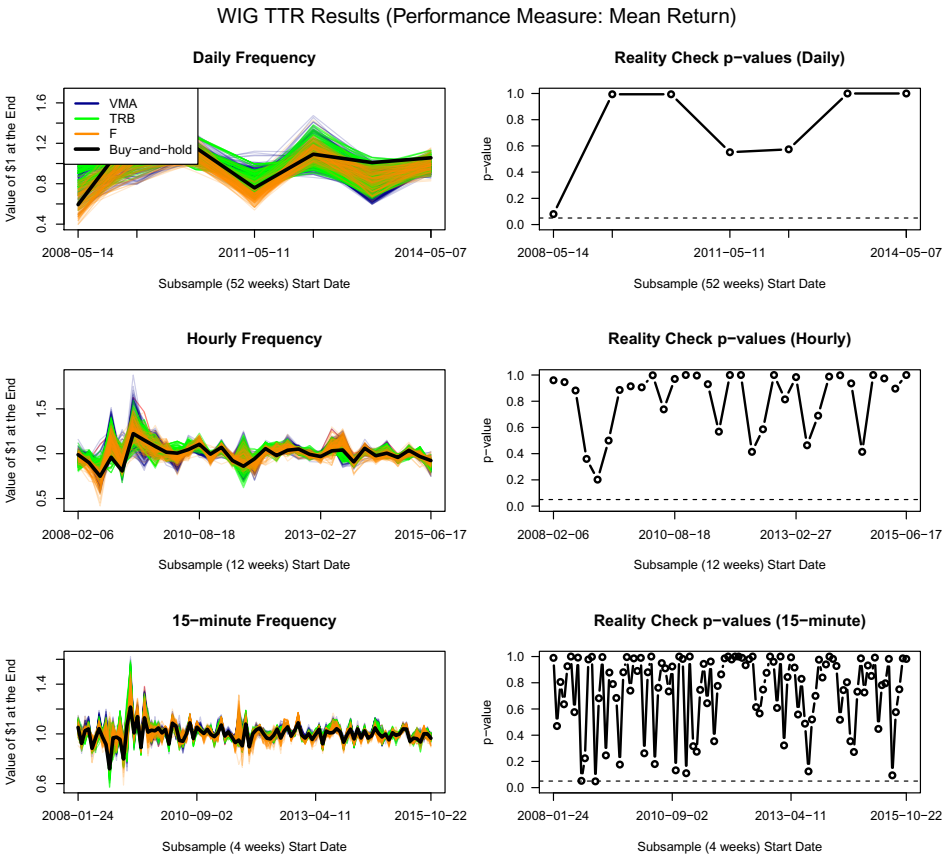


Figure 1.11: Technical Trading Rules evaluation results summary for the Polish index (WIG) and the mean return performance measure.

investment strategy): for each subsample, we identify the corresponding best-performing TTR and denote its index as k^*. In the subsample that follows directly afterwards (indexed by T_i as the beginning and T_j as the end), we follow this particular TTR and record its excess performance, denoted as:

$$\varphi^M_{T_i \to T_j} = \exp\left(\sum_{t=T_i}^{T_j} r_t^{(k^*)}\right) - \exp\left(\sum_{t=T_i}^{T_j} r_t^{(BH)}\right) \tag{1.15}$$

The value $\varphi^M_{T_i \to T_j}$ represents the difference between the terminal value of \$1 invested in the TTR k^* between times T_i and T_j and following the buy-and-hold strategy during the same period. Thus, it may be thought of as the mean return excess performance of the TTR between T_i and T_j.

For a given subsample division $T_0, T_1, \ldots, T_N$, this procedure yields the following time series of values:

$$\varphi^M_{T_1 \to T_2}, \varphi^M_{T_2 \to T_3}, \ldots, \varphi^M_{T_{N-1} \to T_N}. \tag{1.16}$$

Assuming that these values represent a stationary time series, we formulate the null hypothesis that the "Best Rule" investment strategy yields the same expected performance as the buy-and-hold strategy:

$$H_0 : E\left[\varphi^M_{T_i \to T_j}\right] = 0 \tag{1.17}$$

A rejection of H_0 then implies that the expected performances of these two strategies differ in a statistically significant manner. In order to account for a possible autocorrelation in the time series, we estimate the sample mean of the time series (denoted as $\widehat{\mu_\varphi}$) and the standard error of the sample mean as:

$$\begin{aligned} \widehat{\mu_\varphi} &= \frac{1}{N}\sum_{i=1}^{N} \varphi^M_{T_i \to T_{i+1}} \\ SE\left(\widehat{\mu_\varphi}\right) &= \frac{\widehat{\sigma}_\varphi}{N}\left[1 + 2\sum_{m=1}^{N-1}\left(1 - \frac{m}{N}\right)\widehat{\rho}^{(m)}_\varphi\right] \end{aligned} \tag{1.18}$$

where $\widehat{\sigma}$ is the usual standard deviation estimator. The 95% confidence interval for μ_φ is then formed as $\widehat{\mu_\varphi} \pm 1.96 SE\left(\widehat{\mu_\varphi}\right)$. The null hypothesis is rejected when this interval does not contain zero.

We also perform a robustness check of the above procedure. Instead of simply estimating the mean and the standard error that accounts for the possible autocorrelation in the time series, we fit both the $AR(1)$ and $MA(1)$ models to the time series and evaluate the statistical significance of the estimated stochastic process means. Moreover, we perform a non-parametric Wilcoxon signed-rank test to test for the mean value of the time series. All the approaches yield identical results in terms of statistical significance and we omit them for the sake of brevity.

Testing for μ_φ provides an insight into the potential prediction power of the "Best Rule" investment strategy. In order to also evaluate its profitability, for each subsample, we calculate break-even transaction costs. These are the per-trade transaction costs (expressed as the percentage of the traded value) that would wipe out all the excess performance of the TTR. Assuming that the TTR during subsample $T_i \to T_{i+1}$ requires n_i trades, and in each trade, the corresponding one-period return is reduced by TC_i, these transaction costs are break-even if they satisfy the equation:

$$\exp\left(\sum_{t=T_i}^{T_{i+1}} r_t^{(k^*)} - n_i TC_i\right) = \exp\left(\sum_{t=T_i}^{T_{i+1}} r_t^{(BH)}\right). \tag{1.19}$$

Solving for TC_i,

$$TC_i = \frac{\sum_{t=T_i}^{T_{i+1}} r_t^{(k^*)} - \sum_{t=T_i}^{T_{i+1}} r_t^{(BH)}}{n_i}. \tag{1.20}$$

We calculate the break-even transaction costs for each subsample separately (as indexed by i) and for the entire period observed. In this case, the total number of trades is $\sum_{i=1}^{N-1} n_i$ and the entire period break-even transaction costs are equal to:

$$TC_{TOT} = \frac{\sum_{t=T_1}^{T_N} r_t^{(k^*)} - \sum_{t=T_1}^{T_N} r_t^{(BH)}}{\sum_{i=1}^{N-1} n_i} \tag{1.21}$$

In general, the greater the value of TC_i, the greater the evidence of profitability of the trading rule, since in order for the excess performance to disappear, the percentage per-trade transaction costs need to be relatively large. However, we note that, though helpful, this calculation does not provide a definitive answer, since the notion of "relatively large" may be difficult to quantify, especially for the emerging markets considered in our study.

Predictability

The results are displayed in Table 1.4 and, for the 15-minute frequency data, also in graphical form in Figure 1.12. Figure 1.12 displays time development of the excess performance time series for each subsample. Thick horizontal lines represent the estimated mean of these time series (i.e., the estimates $\widehat{\mu_\varphi}$) and colored bars cover the corresponding 95% confidence intervals. If the entire region is above zero, we reject the null hypothesis that the "Best Rule" strategy has zero expected excess performance. These cases are represented by green confidence bands, while for cases that do not reject the hypothesis, the intervals are red.

Summarizing Table 1.4 and Figure 1.12 and assuming that the "Best-Rule" strategy provides a quantifiable measure of the predictable power of technical analysis, we observe that there are qualitative differences in the results, both across stock market indices and across the frequencies considered.

With the exception of the Slovenian index, the "Best-Rule" strategy does not yield excess performance when applied to the daily frequency data. However, we note that the daily frequency results need to be taken with caution, since there are very few observations on which the statistical inference is carried out. Nevertheless, the results are not surprising, since, as argued above, true predictive power of our set of TTRs is not confirmed at the daily frequency. Following the "Best-Rule" strategy at the daily frequency would yield 13% annualized return in excess of the buy-and-hold strategy for the Slovenian index, while applying this strategy to the other indices (with the exception of BUX) would actually yield negative excess returns.

Considering the hourly frequency results, there are four indices (Austrian ATX, German DAX, Slovenian SBITOP, and Polish WIG) for which we document significantly positive values of the "Best-Rule"'s excess performance. These positive results are observed for the CEE countries, with the surprising exception of the marginally significant German DAX index. The results for

the other benchmark index, the SPX, are not significant, and the value $\widehat{\mu_\varphi}$ is very close to zero at all three frequencies.[12] Nevertheless, we observe that moving from a daily into an hourly frequency setting may have improved the performance of technical analysis for some of the stock market indices.

This finding is also consistent with the 15-minute frequency results. In this case, we observe a clear division of indices into two groups. The first group consists of the two well-developed benchmark stock markets (SPX and DAX), for which the values $\widehat{\mu_\varphi}$ are not statistically significant, suggesting that the "Best-Rule" strategy is, on average, no better than simply holding the index. Conversely, the second group consists of the remaining five CEE stock markets. We document a statistically excess performance for each of these markets. These results are in line with our intuition that technical analysis is more likely to possess predictive power for stock markets that are less liquid and relatively less important from the global perspective. Moreover, this predictability is present at higher frequencies to a greater extent than with lower frequencies.

Table 1.4 summarizes these results. Panel A displays estimated values of the per-subsample expected gross[13] excess performance $\widehat{\mu_\varphi}$ of the "Best-Rule" strategy for each of the country-frequency pairs. The values shown in bold are statistically significant in the positive direction at the 5% level. The "Best-Rule" strategy yields a 13% per annum excess return when applied to the daily frequency SBITOP index. At the hourly frequency, the largest excess return is documented for the ATX index: 9.1% per quarter, equivalent to 36.4% per annum. Similarly, at the 15-minute frequency, the ATX index yields as much as 6.1% per month, equivalent to 73.2% per annum.

Profitability - Transaction Costs

Up to this point, the discussion has not taken into account the presence of transaction costs associated with trading. There are several kinds of transaction costs that an investor incurs when opening and closing trading positions (including exchange fees, the costs of immediacy, non-synchronous trading costs, and taxes[14]). We follow an established practice in the literature[15] and calculate the break-even transaction costs, expressed in units of percent-per-trade, that would erase all of the excess performance of the "Best-Rule" strategy. The results are reported in Panel C of Table 1.4. We only report results for the statistically significant predictability cases.

The range of the calculated break-even transaction costs is 2.5 basis points (bps) (WIG

[12]The results for the SPX index are in line with our intuition and thus serve as a good control group for our choice of the entire evaluation setting.

[13]Before transaction costs.

[14]For a detailed discussion, see, for example, Bajgrowicz and Scaillet (2012).

[15]For example, Bessembinder and Chan (1999), Neely and Weller (2003) and Bajgrowicz and Scaillet (2012)

Table 1.4: "Best Rule" investment strategy predictability results.

Panel A reports the estimates of the expected excess performance of the "Best Rule" strategy expressed as the difference between the terminal value of $1 invested in the beginning of a subsample and following either the "Best Rule" strategy or the buy-and-hold strategy. The highlighted values are statistically significant in the positive direction at the 5% level. Panel B reports the same values on an annualized basis. Panel C reports the break-even transaction costs (expressed in percentage points per trade) required to wipe out the excess performance of the "Best Rule" strategy. The values are reported only for the cases in which we possess statistically significant evidence of predictability.

	Panel A: Estimated per-subsample excess performance $\widehat{\mu_\varphi}$			Panel B: Excess performance $\widehat{\mu_\varphi}$ on an annualized basis			Panel C: Break-even transaction costs (per trade) TC_{TOT} (%)		
Frequency	Daily	Hourly	15-Min	Daily	Hourly	15-Min	Daily	Hourly	15-Min
SPX	-0.015	0.008	0.004	-0.015	0.032	0.048			
DAX	-0.091	**0.058**	0.012	-0.091	**0.232**	0.144		0.091	
ATX	-0.007	**0.091**	**0.061**	-0.007	**0.364**	**0.732**		0.115	0.095
BUX	0.069	0.038	**0.026**	0.069	0.152	**0.312**			0.039
PX	-0.025	0.046	**0.029**	-0.025	0.184	**0.348**			0.054
SBITOP	**0.13**	**0.077**	**0.046**	**0.13**	**0.308**	**0.552**	0.758	0.223	0.179
WIG	-0.041	**0.033**	**0.017**	-0.041	**0.132**	**0.204**		0.06	0.025

at the 15-minute frequency) to 75.8 bps (SBITOP at the daily frequency) per trade. The degree to which such transaction costs are likely to actually be present is not quantifiable in a straightforward manner. However, several results may serve as a guideline for determining what reasonable transaction costs might equal. Bajgrowicz and Scaillet (2012) consider the so-called high cost scenario, corresponding to 50 bps per trade and the low cost scenario, corresponding to 25 bps per trade. Neely and Weller (2003) consider 2.5 bps as an estimate of the transaction costs a large institutional investor would face.

Assuming Neely and Weller's value of 2.5 bps, we observe that all statistically significant index-frequency pairs would be profitable even after transaction costs, since in all cases, the break-even transaction costs are greater than 2.5 bps per trade. However, acknowledging that such values may be realistic only for large institutional investors (and likely only for liquid stock markets) and assuming a value of 25 bps instead, the conclusions change considerably, since only the (arguably least-developed) SBITOP index traded at the daily frequency would offer excess returns net of transaction costs. The actual profitability of the "Best-Rule" strategy therefore depends greatly on the particular value of transaction costs an investor would face. Nevertheless, we conclude that our calculated break-even transaction costs are not so low as to unequivocally undermine the documented market-outperforming power of technical analysis.

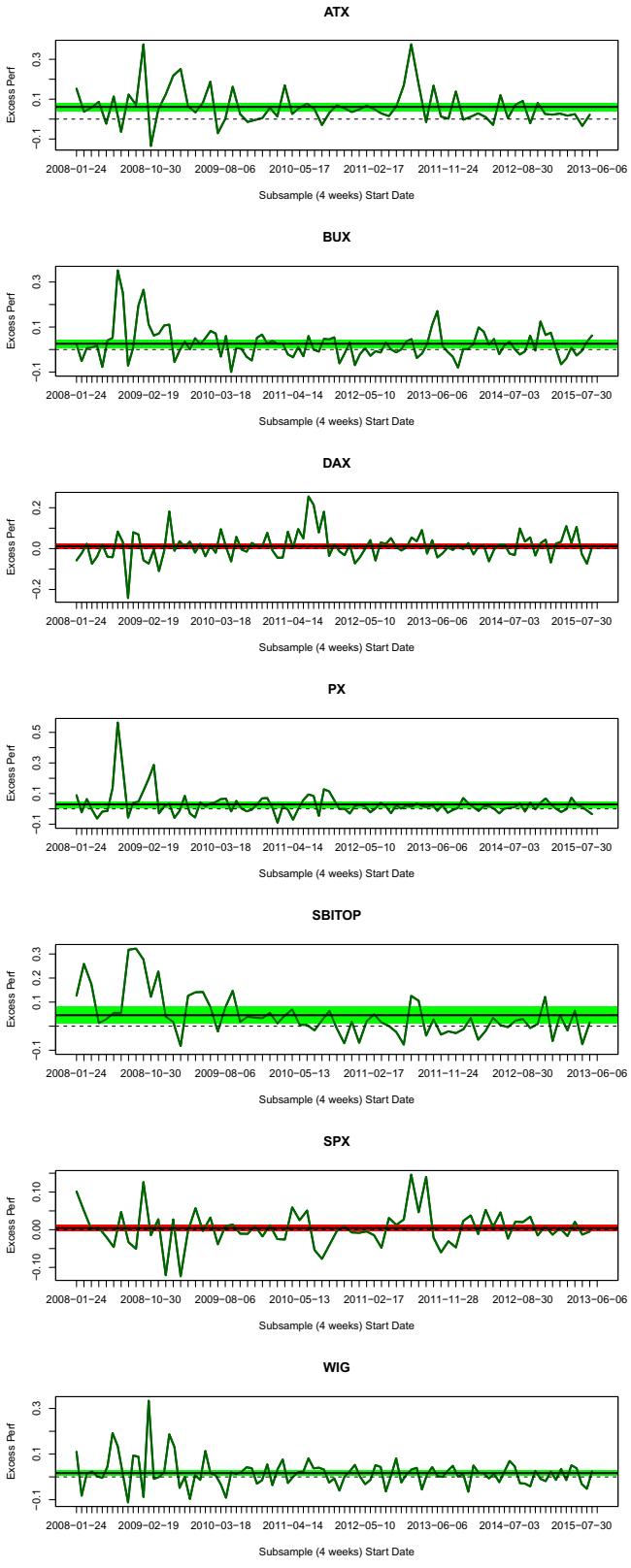


Figure 1.12: "Best Rule" investment strategy graphical results summary for the 15-minute frequency and the mean return performance criterion. The graphs display the time series of excess performance measures $\varphi^{M}_{T_i \to T_{i+1}}$ (green line), their estimated mean value (horizontal thick black line) and the corresponding 95% confidence interval (red and green regions, corresponding to the cases when the null hypothesis is not rejected and is rejected, respectively).

1.6 Conclusion

This study carries out a multidimensional investigation of weak-form market efficiency for the stock market indices of the United States, Germany and five CEE countries during the years following the 2008 financial crisis. Our aim is to analyze the predictive power and profitability of three simple classes of well known technical trading rules applied to market index prices sampled at three different frequencies: daily, hourly, and 15-minute. By dividing our datasets into non-overlapping subsamples, we are able to assess the developments of all results across time. As our sample contains indices representing both highly developed financial markets and markets that can still be considered emerging, we are mainly interested in qualitative comparison of the results across these two groups in the context of the Efficient and Adaptive markets hypotheses.

We approach this research objective from two perspectives. Firstly, we calculate the performance of the set of trading rules on either the mean return basis or the risk-adjusted Sharpe Ratio basis, and evaluate them relative to the benchmark buy-and-hold strategy. In order to avoid data snooping bias, we also conduct White's Reality Check procedure intended to investigate whether the best performing trading rule possesses true predictive power over the price process. Secondly, we study the performance of an investment approach that in each subsample follows the best performing trading rule from the immediately preceding subsample. This "Best-Rule" investment strategy can easily be conducted in the real world, provided an investor is allowed to trade the market index as often as the particular sampling frequency requires. We record the excess performance of this strategy on a subsample basis and test the null hypothesis that the expected excess performance is less than or equal to zero. Additionally, we discuss how a potentially profitable "Best-Rule" strategy persists after accounting for transaction costs.

Combining the results of both approaches, we find that our sample of seven stock market indices can be objectively divided into two groups. The first group, consisting of the US S&P 500 and the German DAX indices exhibits results that strongly point to market efficiency. For the vast majority of subsamples, the Reality Check fails to reject the null hypothesis of no superior performance of the best rule. Moreover, using the best performing rules in the "Best-Rule" investment strategy does not yield excess performance significantly greater than zero. These results are stable over time, and are robust with respect to the choice of the performance measure and the sampling frequency of the price process. These results were largely expected for the two developed stock market indices of the United States and Germany, and we conclude that these indices exhibit results consistent with weak-form efficiency at all three frequencies.

The second group, consisting of the five CEE indices, exhibits similar results only for daily frequency data. For the two intraday frequencies, we find some evidence of time varying informational efficiency of the markets with respect to our set of trading rules. We document several time periods during which White's Reality Check rejects the null hypothesis at one daily (the

Slovenian index) and several intraday frequencies (all three indices). Moreover, an investor following the "Best Rule" investment strategy would obtain statistically significant positive excess returns for the Austrian, Slovenian, and Polish indices at the hourly frequency and for all five indices at the 15-minute frequency. However, these results are challenged after the calculated breakeven transaction costs (varying from 2.5 bps to 75.8 bps) are taken into account, although we do not rule out the possibility that the net excess returns would also remain positive. Overall, we conclude that the results for this second set of indices suggest an adaptive nature of market efficiency, with the occurrence of periods during which technical analysis might be truly profitable.

We conclude that this study highlights differences between well-developed and emerging stock markets. Since analysis of informational efficiency is of great importance, especially for the financial markets that can still be considered to be emerging, two possible directions for future research may include a detailed analysis of transaction costs and an investigation of additional trading strategies as well as of even higher frequency trading profitability.

1.A Appendix

1.A.1 Trading Rules

Three different classes of technical trading rules (TTRs) are employed in this study: the Variable Moving Average (VMA), the Trading Range Breakout (TRB), and the Filter rule (F).

1. **Variable Moving Average ($VMA(N_S, N_L, B)$):**

   - $N_S \in \mathbb{N}$: number of past closing prices used to calculate the short moving average,
   - $N_L \in \mathbb{N}$: number of past closing prices used to calculate the long moving average, $N_L > N_S$,
   - $B \in (0, 1)$: the percentage band of inactivity.

   The trading rule works as follows. Denote by P_t the closing price at time t and by S_t the trading signal for period t. Start with a neutral position, thus, $S_0 = 1$. At each trading period t, the two moving averages of past prices are calculated:

$$SMA_t = \frac{1}{N_S} \sum_{i=t-N_S}^{t-1} P_i, \quad LMA_t = \frac{1}{N_L} \sum_{i=t-N_L}^{t-1} P_i.$$

   Based on this information, the trading signal for period t, S_t, is generated based on the following instructions:

- If $S_{t-1} \in \{-1, 0\}$ and $SMA_t \geq (1+B)LMA_t$, then $S_t = 1$.
- If $S_{t-1} = 1$ and $SMA_t \geq LMA_t$, then $S_t = 1$.
- If $S_{t-1} \in \{0, 1\}$ and $SMA_t \leq (1-B)LMA_t$, then $S_t = -1$.
- If $S_{t-1} = -1$ and $SMA_t \leq LMA_t$, then $S_t = -1$.

Effectively, starting in a neutral position, the first time the short moving average rises above (falls below) the long moving average adjusted by the band of inactivity B, a long (short) position is entered. This long (short) position is maintained until the short moving average stays above (below) the long moving average value (unadjusted by B). As soon as this condition is violated, a neutral position is taken. This position is maintained until the short moving average rises above (falls below) the long moving average adjusted by the band of inactivity B, at which point a long (short) position is entered again.

We set the following values of the parameters:

$$N_S \in \{2, 3, 5, 10, 15\}, N_L \in \{10, 15, 20, 25, 30, 40, 50, 75\},$$

$$B \in \{0, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007\}$$

Recognizing that for each rule, it must hold that $N_S < N_L$, this parametrization yields in total 296 individual TTRs.

2. **Trading Range Breakout ($TRB(N, B)$):**

- $N \in \mathbb{N}$: number of past closing prices used to evaluate the trading range,
- $B \in (0, 1)$: the percentage band of inactivity.

We again start in a neutral position, $S_0 = 0$. At each time period t, the following two values (the resistance and the support, respectively) are calculated:

$$\alpha_t = \max_{i \in \{t-N, \ldots, t-2\}} P_i, \quad \beta_t = \min_{i \in \{t-N, \ldots, t-2\}} P_i.$$

Based on this information, the trading signal for period t is generated based on the following instructions:

- If $S_{t-1} = 0$ and $P_{t-1} > (1+B)\alpha_t$, then $S_t = 1$.
- If $S_{t-1} = 0$ and $P_{t-1} < (1-B)\beta_t$, then $S_t = -1$.
- If $S_{t-1} = 1$ and $P_{t-1} < (1+B)\beta_t$, then $S_t = 0$.
- If $S_{t-1} = -1$ and $P_{t-1} > (1-B)\alpha_t$, then $S_t = 0$.

- Else, $S_t = S_{t-1}$.

Effectively, starting in a neutral position, the first time the closing price rises above (falls below) the maximum (minimum) of the past $N-1$ closing prices, adjusted by the band of inactivity B, a long (short) position is entered. This long (short) position is maintained until the closing price gets below a critical level defined as the minimum (maximum) of past N prices, adjusted by the band of inactivity B. This band of inactivity serves now as a tolerance level in a sense that in the case of a long position, the closing level is obtained by multiplying the minimum of the past N prices by $(1+B)$. As soon as the price gets below (rises above) this closing level, a neutral position is taken. This position is maintained until the closing price rises above (falls below) the maximum (minimum) of the past $N-1$ closing prices, adjusted by the band of inactivity B, at which point a long (short) position is entered again.

We set the following values of the parameters:

$$N \in \{5, 7, 10, 12, 14, 16, 18, 20, 22, 25, 30, 35, 40, 45, 50, 75\},$$

$$B \in \{0, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01\}$$

This parametrization yields in total 176 individual TTRs.

3. **Filter ($F(x, y)$):**

- $x > 0$: the percentage filter for entering a position,
- $y > 0$: the percentage filter for exiting a position.

We again start in a neutral position ($S_0 = 0$) and we define the reference price as $R = P_0$. At each subsequent time period t, the trading signal for period t is generated based on the following instructions:

- If $S_{t-1} = 0$ and $P_{t-1} > (1+x)R$, then $S_t = 1$ and define $R = P_{t-1}$.
- If $S_{t-1} = 0$ and $P_{t-1} < (1-x)R$, then $S_t = -1$ and define $R = P_{t-1}$.
- If $S_{t-1} = 1$ and $P_{t-1} < (1-y)R$, then $S_t = 0$ and define $R = P_{t-1}$.
- If $S_{t-1} = 1$ and $P_{t-1} > R$, then $S_t = 1$ and define $R = P_{t-1}$.
- If $S_{t-1} = 1$ and $P_{t-1} \in [(1-y)R, R]$, then $S_t = 1$.
- If $S_{t-1} = -1$ and $P_{t-1} > (1+y)R$, then $S_t = 0$ and define $R = P_{t-1}$.
- If $S_{t-1} = -1$ and $P_{t-1} < R$, then $S_t = -1$ and define $R = P_{t-1}$.
- If $S_{t-1} = -1$ and $P_{t-1} \in [R, (1+y)R]$, then $S_t = -1$.

Effectively, starting in a neutral position, the first time the closing price rises above (falls below) the reference price R adjusted for the opening filter x, a long (short) position is entered. At the same time, a new reference price is defined as the closing price at which the position was entered. If during this long (short) position the next period price increases (decreases), the reference price is updated as this new closing price and the long (short) position is maintained until the closing price falls below (rises above) the closing-filter-adjusted reference price $(1+y)R$ $((1-y)R)$, at which point a neutral position is taken and the reference price R is again re-defined as this last closing price. This neutral position is maintained until the closing price rises above (falls below) the level $(1+y)R$ $((1-y)R)$, at which point a long (short) position is entered again.

We set the following values of the parameters:

$$x \in \{0.0025, 0.005, 0.0075, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05\},$$

$$y \in \{0.0025, 0.005, 0.0075, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05\}$$

This parametrization yields in total 144 individual TTRs.

Our complete "universe" of trading rules thus contains $296 + 176 + 144 = 616$ individual TTRs.

1.A.2 Reality Check Stationary Bootstrap

Stationary bootstrap of Politis and Romano (1994) is used to obtain the asymptotic distribution of the White (2000) Reality Check test statistic V valid under the null hypothesis. We start with the original time series of prices $\{P_1, \ldots, P_T\}$. One stationary bootstrap sample is defined by the sequence of time indices $\theta(t) \in \{1, \ldots, T\}$ $\forall t$. For a given value of the algorithm parameter $q \in (0, 1)$, this sequence is generated as follows:

1. Set $t = 1$. Draw $\theta(t) = \theta(1)$ independently and uniformly from the set {1, ..., T}.
2. Set $t = t + 1$. If $t > T$, stop. Otherwise, draw one random independent realization from the standard uniform distribution $U[0, 1]$ and denote this number by u.
   a) If $u < q$, draw $\theta(t)$ independently and uniformly from the set {1, ..., T}.
   b) If $u \geq q$, set $\theta(t) = \theta(t-1) + 1$. If $\theta(t) > T$, reset $\theta(t) = 1$.
3. Repeat Step 2.

The bootstrapped time series of prices is then given by $\{P_{\theta(1)}, P_{\theta(2)}, \ldots, P_{\theta(T)}\}$. Effectively, the algorithm samples blocks of original data of varying length, where the length of one block is a random variable generated from a discrete geometric probability distribution with the (success)

parameter q. An expected length of one block is therefore $1/q$. We follow a standard practice in the literature (Sullivan, Timmermann, and White (1999), Park et al. (2014)) and set $q = 0.1$.

For this bootstrapped sample, a universe of TTRs is applied and a bootstrapped value of the test statistic is calculated as:

$$V_b^* = \max_{k=1,\ldots,K} \{\sqrt{T}\left(\hat{f}_{M,b}^{(k)*} - \hat{f}_M^{(k)}\right)\}.$$

This completes one bootstrap repetition of the test. This procedure is repeated B times and the empirical distribution of V_b^* approximates the probability distribution of the test statistic V valid under the null hypothesis of no abnormal performance.

1.A.3 Sharpe Ratio Performance Measure Results

The next seven figures (Figures 1.13 to 1.19) display the technical trading rules performance evaluation for the Sharpe Ratio criterion, which we omitted in the main body of the text for the sake of brevity. We emphasize that overall, our entire analysis appears to be highly robust with respect to the particular choice of performance measure. Thus, if a rule is deemed to bear true outperforming power over the market, the returns emanating from this rule outperform the buy-and-hold returns also from the risk-adjusted perspective.

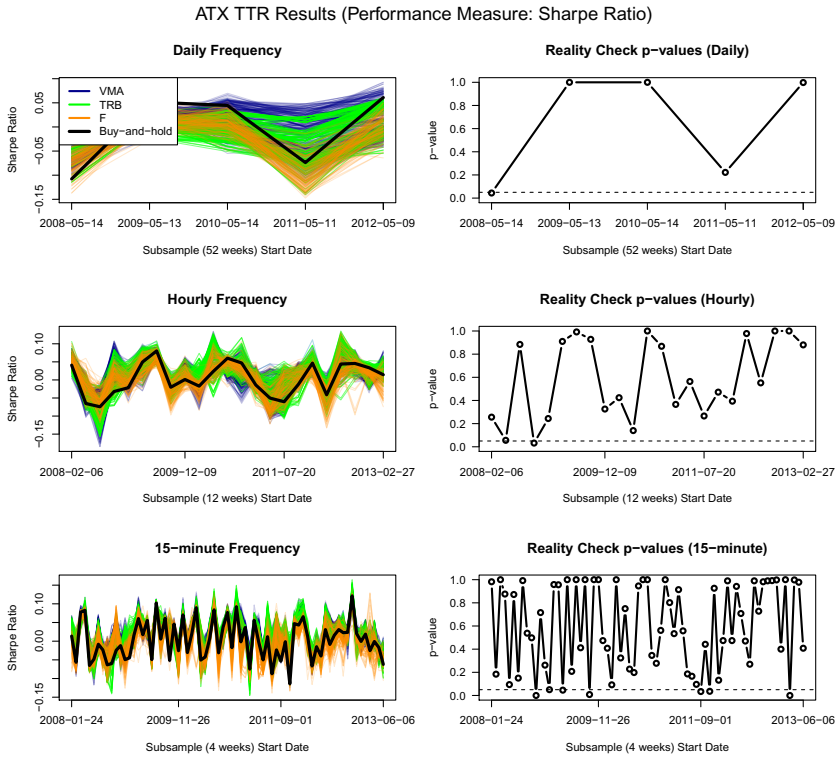


Figure 1.13: Technical Trading Rules evaluation results summary for the Austrian index (ATX) and the Sharpe Ratio performance measure.

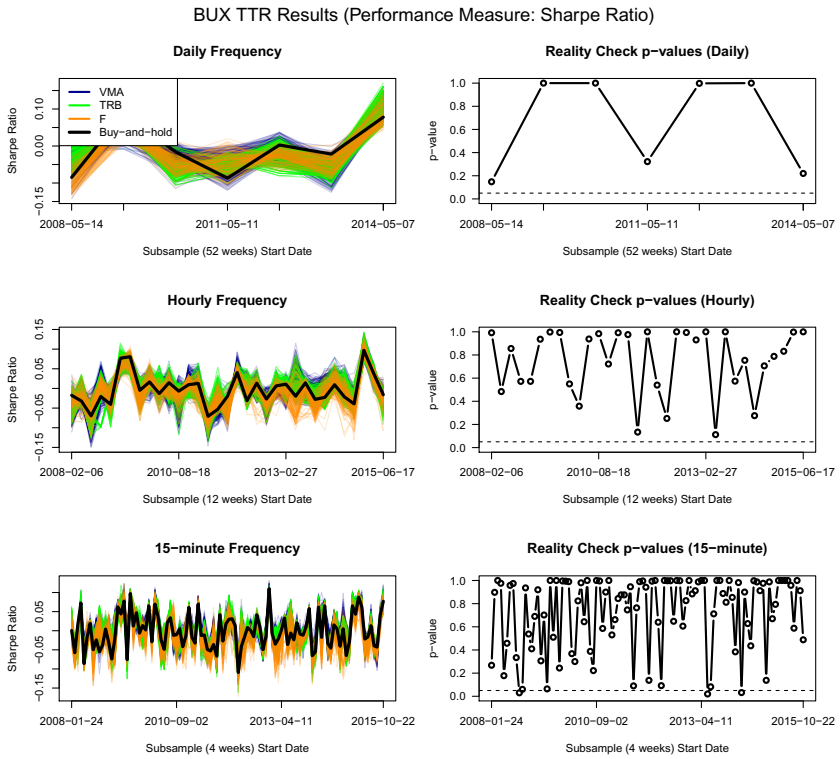


Figure 1.14: Technical Trading Rules evaluation results summary for the Hungarian index (BUX) and the Sharpe Ratio performance measure.

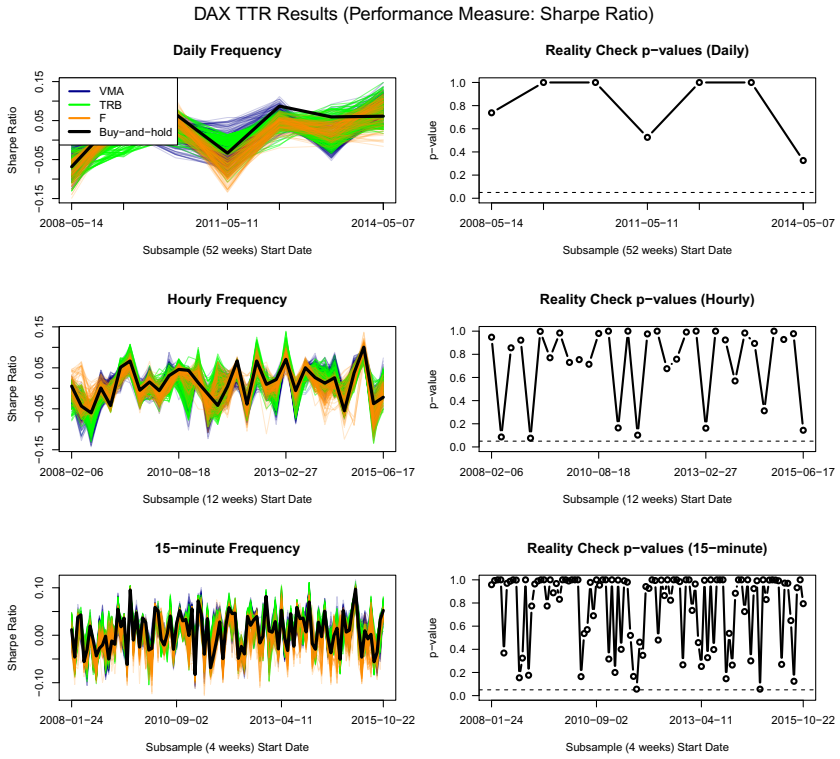


Figure 1.15: Technical Trading Rules evaluation results summary for the German index (DAX) and the Sharpe Ratio performance measure.

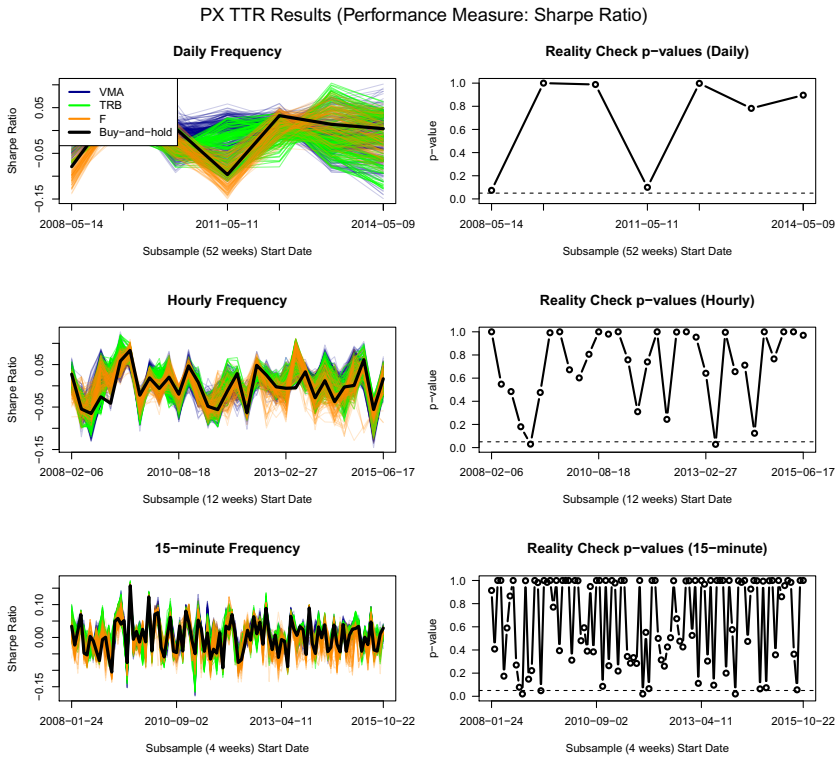


Figure 1.16: Technical Trading Rules evaluation results summary for the Czech index (PX) and the Sharpe Ratio performance measure.

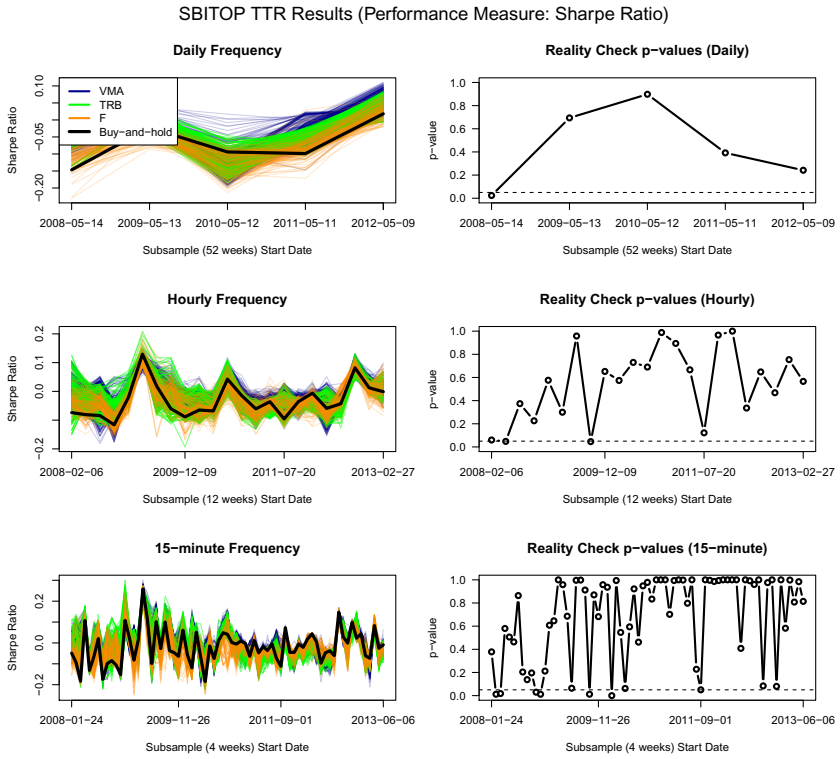


Figure 1.17: Technical Trading Rules evaluation results summary for the Slovenian index (SBITOP) and the Sharpe Ratio performance measure.

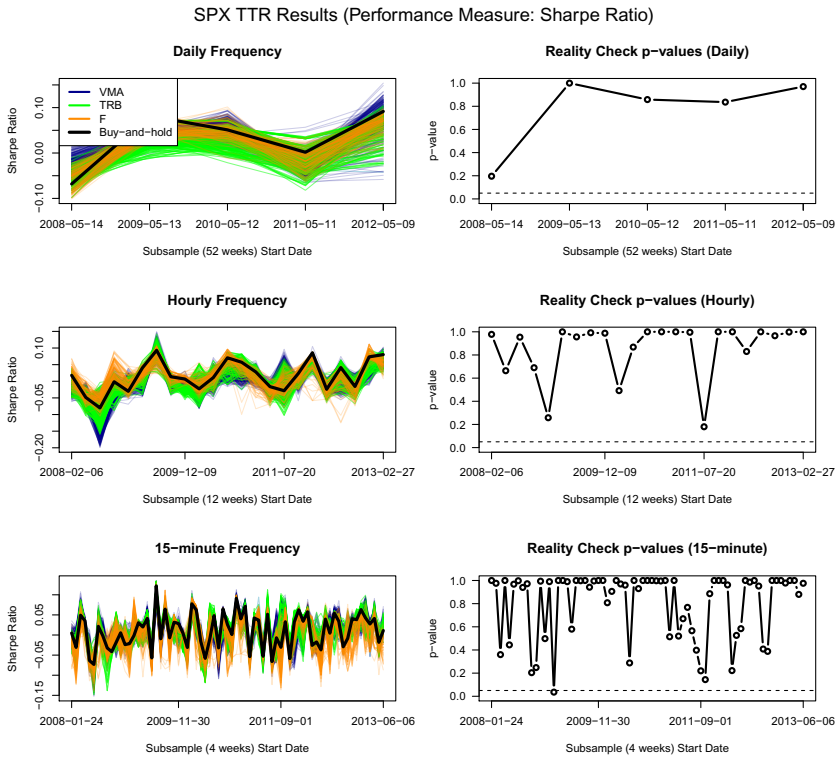


Figure 1.18: Technical Trading Rules evaluation results summary for the S&P 500 index (SPX) and the Sharpe Ratio performance measure.

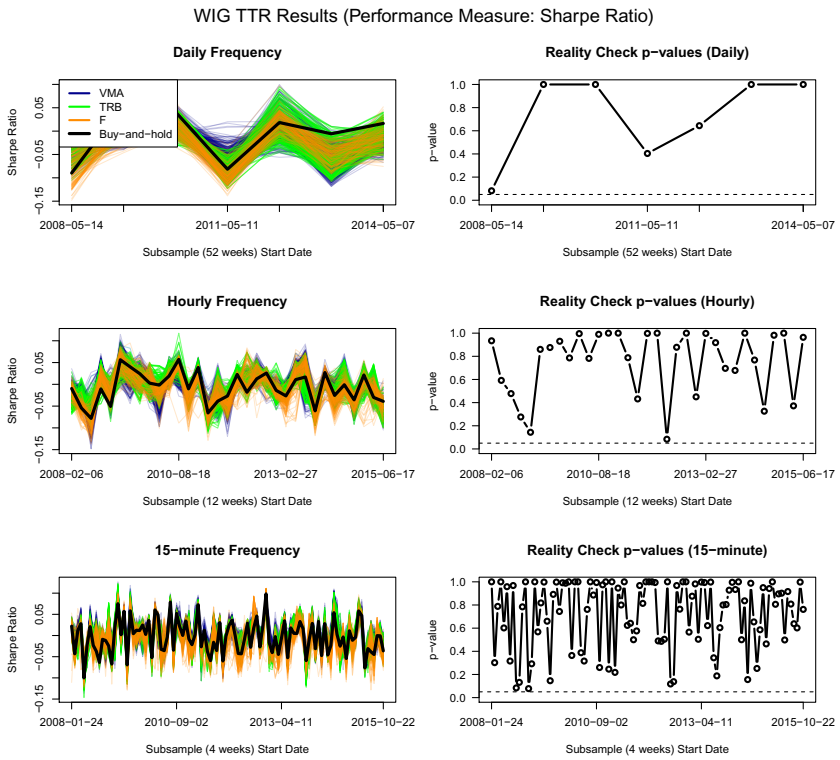


Figure 1.19: Technical Trading Rules evaluation results summary for the Polish index (WIG) and the Sharpe Ratio performance measure.

Chapter 2

Order Flow Toxicity Around Oil Inventory Announcements: Evidence of the VPIN Metric.

2.1 Introduction

One of the main research areas in the market microstructure literature is the relationship between market makers and market takers, that is, between the passive and the active (aggressor) sides of each trade in the limit order book market. Traditionally, market makers have been assumed to be competitive and to face the problem of adverse selection, referring to the possibility that the aggressive market taker possesses private information about the asset value, and that the bid and ask prices quoted by market makers may be exploited by such an informed trader, leading to market maker's losses. Consequently, an accepted theoretical result is that, in an effort to contain the risk of order flow toxicity, market makers' reaction is to adjust their bid-ask spread accordingly. Greater probability of market makers being adversely selected may therefore cause wider bid-ask spreads, smaller market depth (the number of contracts passively quoted at the best bid and offer prices), and an increase in the price impact of trades (Glosten and Milgrom (1985), Easley and O'Hara (1987)). The relationship between the degree of adverse selection and various limit order book market characteristics is, therefore, naturally of interest to all market participants and regulatory authorities, as is the availability of appropriate methodology by which the presence of informed trading can be quantified in real time.

In their seminal paper, Easley et al. (1996) develop a sequential trading model, in which competitive market makers use a Bayesian updating technique to protect themselves from the order flow toxicity imposed by informed traders. The measure of the adverse selection proposed

by this model, the PIN (the Probability of Informed Trading), is based on order book side imbalance of the total volume traded, while estimation of the model parameters is carried out using maximum likelihood. Using transaction level data of multiple stocks listed on the NYSE, the authors document a positive relationship between PIN and quoted bid-ask spreads.[1]

The Easley et al. (1996) model became the cornerstone of empirical research into the presence of adverse selection in limit order book markets and has since been applied to a wide range of settings (Easley, Hvidkjaer, and O'Hara (2002), Vega (2006), Chen and Zhao (2012) and Chang and Lin (2015)). Of course, the methodology itself has also been extended and modified, as the static nature of the original model carries several limitations, especially when applied to modern and fast electronic markets. In particular, as Weng et al. (2017) assert, in the original model, the rates of arrival of informed and uninformed traders are assumed to be constant over the observed period (which, according to the original assumptions, represents several days to several months worth of data), making a timely identification of the presence of short lived information virtually impossible.

With the purpose of addressing the limitations of the static parametric PIN model, Easley, de Prado, and O'Hara (2010) introduced a non-parametric dynamic way to estimate the informed trading probability. Their measure, called the Volume-synchronized Probability of Informed Trading (VPIN), enables a researcher to update the estimated degree of adverse selection present in the market on a sequential basis. This sequentiality is based on a volume-time based approach of the VPIN metric, producing a new estimate of VPIN after each pre-determined number of contracts is traded in the market (an approach referred to by the authors as volume bucketing). A particular feature of volume bucketing is that the rate of VPIN updating is synchronized with trading intensity, which the authors consider to be the proxy for the rate of the arrival of new information. Therefore, assuming that the calculation parameters are chosen appropriately, a researcher may possess relatively frequent updates of VPIN, creating an attractive empirical research area, especially for today's highly liquid, competitive, and fast-moving markets.[2]

Despite numerous studies investigating the VPIN metric in various settings (for example, Abad and Yagüe (2012), Wei, Gerace, and Frino (2013), (Wu et al. 2013), Cheung, Chou, and Lei (2015), Lee et al. (2017) and Yildiz, Van Ness, and Van Ness (2020)), there is an ongoing debate regarding its validity in truly capturing the order flow toxicity. Moreover, although the static PIN and its modified dynamic version have been used in event studies (Vega (2006), Aktas et al. (2007), Chen and Chung (2007)), to the best of our knowledge, the number of studies investigating the VPIN dynamics around regularly occurring macroeconomic announcement events remains relatively low, particularly so in the environment of high frequency limit order

[1] A more detailed description of this model is presented in section 2.4.
[2] A more detailed description of this model is presented in section 2.4.

book markets.[3]

The aim of this study is to help to fill this gap by investigating the dynamics and the predictive ability of VPIN during regular, relatively frequent, and significant macroeconomic announcement events in modern electronic markets. Our asset of choice is one of the most liquid and frequently traded securities in the world - the WTI crude oil futures contracts traded on NYMEX.[4] One of the distinct features of the crude oil futures market is the presence of regular and significant macroeconomic data announcements that directly affect the price of crude oil: the weekly data on the level of US crude oil inventories. In particular, two significant data announcements are released each week: the Weekly Statistical Bulletin by the American Petroleum Institute (henceforth, the API report) and the Weekly Petroleum Status Report by the Energy Information Association of the Department of Energy of the United States (henceforth, the DOE report).

The API and the DOE reports, though aiming to report the same quantities, differ in the ways, the means, and the authority by which they collect the underlying data, often resulting in different reported values of crude oil inventory levels. While the API is a private organization, and companies involved in the supply chain of crude oil are not required by law to disclose information to the API, the reverse is true for the US Government's DOE report. Moreover, while the API report is usually released each Tuesday after market hours (at 4:30 p.m. Eastern Time (ET)), the DOE report is usually released the following Wednesday during market hours (at 10:30 a.m. ET). This unique and regular sequential publishing schedule of the two reports, combined with the high frequency trade-level data on the WTI futures contract at our disposal, provide an interesting setting in which the changes in various market microstructure characteristics (including the estimates of order flow toxicity) occurring at very high frequencies can be investigated.

We are particularly interested in estimating the dynamics of the order flow toxicity (as measured by the VPIN metric) around the API and DOE announcements, and its dependence on the actual quantities announced, and on the degree to which the quantities differ from market expectations. We measure these market expectations as the DOE report's survey forecast (available prior to the release of the API report) for the API report, and as the published API value for the DOE report.[5] We aim to contribute to the academic debate on VPIN and the adverse selection measurement by answering the following questions: Is the presence of the announcements accompanied by changes in the VPIN dynamics? Do features exist in the VPIN dynamics that depend on the degree to which the announcements are surprising? Do the estimated VPIN

[3]A more detailed overview of similar studies is provided in the Literature Review, section 2.2.

[4]See, for example, `https://www.cmegroup.com/education/featured-reports/cme-group-leading-products.html`

[5]A similar approach is taken by Ye and Karali (2016), Armstrong, Cardella, and Sabah (2017) and Anatolyev, Seleznev, and Selezneva (2018).

values suggest a potential presence of insider trading? Are the answers to the previous questions different for the two types of the announcements? Finally, how do the VPIN calculation input parameters affect the observed results?

We find that the VPIN dynamics observed around oil inventory announcements are in accordance with the hypothesized presence of greater informed order flow intensity immediately after new information is published. In particular, the estimated probability of informed trading increases by approximately 20% on average for both reports under investigation. The recorded VPIN adjustment, however, occurs notably faster for the DOE report, which is usually released during the pit trading hours and is accompanied by considerably greater overall trading activity. Moreover, we find some evidence of a positive association between the magnitude of the VPIN change around the two types of inventory announcements and the degree to which these announcements are surprising. However, we note that changes in the informed order flow dynamics are observed only after publication, as we find no evidence of an association between the pre-announcement VPIN dynamics and the degree to which an announcement is surprising. Because one of the possible reasons for such rapid changes in informed trading activity could be the presence of inside information (especially of the last minute leakage kind), the fact that we do not observe such dynamics prior to oil inventory announcements does not provide evidence of the presence of insider trading.

Interestingly, on DOE announcement days, we observe, on average, 20% smaller values of VPIN during the 30 minutes leading up to the publication time, as opposed to the same time of day on non-announcement days. This average observed difference of approximately 3 percentage points is, however, not observed, when we consider the corresponding time windows pertaining to the API report. We conjecture that this observation may be associated with relatively increased activity of liquidity-demanding traders aiming to reduce any oil price exposure they might face due to an upcoming announcement, as well as with relatively decreased activity on the part of informed and speculative traders, who are less willing to engage in usual trading activities, again, because of a pending announcement.

Finally, although differing in the absolute terms (the average non-announcement day VPIN is equal to 0.175 for the smaller volume bucket size category and 0.125 for the larger bucket size category), the qualitative nature of the changes observed in VPIN dynamics appears to be robust with respect to the researcher's choice of the volume bucket size, which represents one of the key VPIN metric input parameters. We thus assert that VPIN may, under certain circumstances, be considered a viable and valuable measure of order flow toxicity, applicable even in electronic markets with great order flow intensity.

The remainder of this paper is structured as follows. Section 2.2 overviews studies focusing either on the phenomenon of adverse selection or on crude oil inventory announcements. Section 2.3 describes our three-year sample of WTI futures contracts trade data and the complemen-

tary data on the API and DOE reports. In section 2.4, we introduce both the original PIN model and its younger semi-parametric sibling, the VPIN. Further, we provide definitions of all variables used in hypothesis testing, and describe the empirical strategy we use to evaluate our hypotheses. Finally, section 2.5 evaluates the dynamics of various market characteristics during hours surrounding the publication of the inventory reports, the dynamics of the VPIN metric, and its relationship with the undelying macroeconomic fundamentals in the form of published oil inventory data.

2.2 Literature Review

2.2.1 Informed Trading, Adverse Selection and VPIN in High-Frequency Markets

In the Easley et al. (1996) sequential model, both informed and uninformed traders are present and, at the beginning of each period, an information event impacting the asset value occurs (good news, bad news, and no news are the three possibilities considered). The aggressors then arrive according to the Poisson distribution, while the intensity parameter of this distribution depends on types of traders as well as on the occurrence of the information event.

As mentioned earlier, the original static model has been extended multiple times, most notably by Easley et al. (2008), in which the authors use a GARCH-type model to assume time varying rates of arrival of different classes of traders, while retaining the parametric nature of the model. Studies investigating this dynamic model include Engle and Neri (2010) and Agudelo, Giraldo, and Villarraga (2015).

Finally, a non-parametric VPIN model was introduced by Easley, de Prado, and O'Hara (2010). The authors demonstrate the usefulness of the VPIN metric on the E-mini S&P 500 and the WTI crude oil futures contracts, sampled during the period from January 2008 to June 2011. The authors find that the VPIN metric tends to precede a large fraction of periods of extreme volatility, especially those induced by order flow toxicity accompanied by large risks of adverse selection. Moreover, the VPIN metric may be useful to distinguish the causes of volatility. Although mentioning one instance of the DOE report announcement (December 9, 2009), the authors do not systematically evaluate the relationship between VPIN and the weekly crude oil inventories information events represented by the API and DOE reports.

Easley, de Prado, and O'Hara (2011) focus specifically on the dramatic episode that occurred on May 6, 2010 (also known as the 2010 Flash Crash), during which various major US stock market indices experienced a rapid collapse followed by a quick and substantial recovery.[6] By the authors' calculations, changes in the VPIN metric preceded the dramatic changes in the

[6]For example, the DJIA index experienced its largest one-day drop to date, falling by 998.5 points.

E-mini S&P 500 futures contract. Moreover, emphasizing the usefulness of VPIN, the authors argue that other market volatility measures (such as the VIX index) did not reach levels as extreme as the VPIN metric soon enough to provide a timely indication of such extreme order flow toxicity.

Multiple studies investigate the behaviour of VPIN in various settings. Abad and Yagüe (2012) calculate VPIN for Spanish stock market data. The authors find that the two input parameters of the VPIN metric provide reasonable flexibility that has the ability to allow a researcher to extend the VPIN applications beyond the HFT world.

Wei, Gerace, and Frino (2013) apply the VPIN metric to a set of securities traded at the Australian Stock Exchange during the period from January 2008 to December 2010. They examine the causal relationship between VPIN and several market characteristics (quote imbalance, price volatility and duration), asserting that statistically significant Granger causality in the direction from VPIN towards these factors is observable. However, Granger causality in the reverse direction is observable as well, leading the authors to conclude that there exists a dynamic two-way relationship between the variables.

Wu et al. (2013) investigate the behaviour of various indicators of market microstructure dynamics and conclude that VPIN does indeed provide valuable information regarding approaching periods of market distress. Utilizing a very large dataset (tick data on 93 highly liquid futures contracts spanning a period of 67 months, including the WTI futures), and the processing power of the United States' second most powerful supercomputer, the authors conduct a deep VPIN investigation, including on its sensitivity to the choice of the underlying parameters. VPIN's strong predictive power to anticipate liquidity-induced volatility has been documented. Similar conclusions regarding the 2010 Flash Crash episode were achieved by Menkveld and Yueshen (2013).

Cheung, Chou, and Lei (2015) examine the information content of VPIN around the call events of callable barrier option contracts traded on the Hong Kong exchange spanning 2008 and 2009. Using the bulk classification algorithm proposed by Easley, de Prado, and O'Hara (2012), the authors document a considerable increase in VPIN around these call events, concluding that there is a greater market risk surrounding them.

We must note, however, that the VPIN metric considered in this paper has not been uniformly accepted in the academic literature. In particular, an interesting dispute between the VPIN authors and Andersen and Bondarenko (2014b), has occurred. Contrary to the original authors' claims, Andersen and Bondarenko (2014b) fail to document significant predictive ability of VPIN in detecting short-term price volatility. The authors argue that the main source of seemingly strong predictive power is the strong correlation between VPIN and the underlying trading intensity, meaning that VPIN has no incremental predictive ability.

Easley, de Prado, and O'Hara (2014) react to these conclusions, stating that "[Andersen and

Bondarenko] attack a methodology we do not advocate, an analysis we never performed, and conclusions we did not draw." Finally, Andersen and Bondarenko (2014a) review the reaction's claims and collect additional empirical evidence of VPIN being prone to distortions driven by trading intensity and price volatility. Moreover, the authors document that the VPIN values are sensitive to the choice of the underlying trade classification technique, the use of which is necessary whenever the data on the aggressor side of each trade is not available. Due to the fact that the dataset in our study contains such aggressor side information, these sensitivity issues are not present and the VPIN results applied to our data are not susceptible to distortions of this kind.

An alternative to the VPIN was proposed by Chang, Chang, and Wang (2014). The authors assert that their measure of intraday informed trading, called the Dynamic Probability of Informed Trading (DPIN), is especially well-suited for estimation of order flow toxicity at finer frequencies, such as 15 minutes. Applying the DPIN metric to the NYSE listed stocks transaction-level data, they document a statistically significant relationship between the degree of private information and firm-specific return variation. We note, however, that since our aim is to investigate informed trading dynamics around very informative and liquid announcement events, constraining the calculation frequency to 15 minutes might not yield informed trading estimates with satisfactory granularity. The ability of VPIN to adjust the rate of updating to the actual trading intensity represents a significant advantage of the VPIN metric in our setting.

Weng et al. (2017) utilize Taiwanese stock market index futures data during the period from January 2003 to December 2007 and compare the DPIN and VPIN metrics, focusing particularly on both foreign and domestic institutional investors' informed trading participation and comparison thereof. Lee et al. (2017) apply the VPIN metric to a unique dataset containing the information on the locations of institutional investors trading US futures contracts during the 2008-2009 financial crisis. The authors document a significant positive impact of foreign institutional investors' VPIN on the assets' returns. Interestingly, a significant impact of domestic institutional investors' VPIN is observed solely on Wednesdays.

Most recently, Yildiz, Van Ness, and Van Ness (2020) study the real-time risk management potential of the VPIN metric by analyzing moments of sudden liquidity deteriorations in US equity markets. They document ex-ante predictive ability of VPIN in forecasting market liquidity and volatility, leading them to suggest that VPIN indicates the adverse selection problem that liquidity providers face by capturing the valuable information in trade volume data.

2.2.2 Crude Oil Inventories

The relationship between energy commodities inventories announcements and corresponding market price formation characteristics has been extensively studied in the academic literature.

Linn and Zhu (2004) focus on natural gas futures prices and weekly natural gas storage change reports by the Energy Information Association (EIA), documenting a significant increase in intra-day volatility during announcement days. Chang, Daouk, and Wang (2009) analyze Bloomberg professionals' forecasts of inventory level changes and conclude that crude oil futures returns respond to inventory announcements in an immediate manner. Moreover, the magnitude of the response is greater if the announced value is unexpected, and even more so if the forecasts were produced by analysts with good track records.

More recently, Bu (2014) utilizes a GARCH model to investigate the relationship between the degree of surprise of a DOE report and the magnitude and volatility of crude oil returns. While a negative relationship between inventory shocks and returns is documented, no dynamics are observable between the inventory shocks and the announcement day volatility of oil prices.

Bjursell, Gentle, and Wang (2015) focus on the identification of price jumps in three energy futures markets (crude oil, heating oil, and natural gas) during the period from 1990 to 2008, focusing particularly on periods surrounding the corresponding inventory announcements. Applying several non-parametric price jump identification methods, the authors document that the inventory announcement dates are indeed accompanied by greater price jump components of the price discovery process. Interestingly, post-announcement volatility tends to return to a normal level more quickly during announcements with price jumps, as opposed to announcements not accompanied by price jumps.

Ye and Karali (2016) evaluate the price and volatility reactions of the WTI crude oil futures market (sampled at the 5-minute frequency) to both API and DOE reports. The API surprise is defined as the difference between the API report and the Reuters forecast by professional analysts, while the DOE surprise is defined as the difference between the API and the DOE reports, that is, the API report value is assumed to be the expected value of the DOE report (in this case, the same logic is adopted in this study). The authors find that both reports instigate an immediate and inverse shock to the price of crude oil, observing generally stronger effects for the DOE report. Moreover, they document a relationship between the size of price changes due to the DOE surprise and the direction and the strength of the API surprise itself.

Our study is perhaps most closely related to papers by Armstrong, Cardella, and Sabah (2017) and Bjursell, Wang, and Zheng (2017). Armstrong, Cardella, and Sabah (2017) evaluate the impact of the weekly DOE report on various oil market liquidity measures. These measures include the number of trades, the quoted bid-ask spread, the depth of the limit order book, and the absolute order imbalance. Using WTI futures data during the period from January 2010 to December 2015, the authors find that the publication of the DOE report very quickly moves prices in the direction of the new information and subsequently leads to increased liquidity (as measured by the above characteristics). The authors argue that, rather than adverse selection, these measures capture the degree of information flow into prices and they advise caution when

interpreting this information flow as the degree of informed trading.

Bjursell, Wang, and Zheng (2017) consider the crude oil and natural gas futures markets during the 2009 - 2015 period. Utilizing trade-level data (not containing aggressor-side information) and considering the announcements by the Energy Information Association only, they document significant increases in the VPIN metric for the announcements accompanied by price jumps, as well as for the unexpected price jumps in these futures markets. The authors observe greater order flow toxicity as measured by VPIN for the natural gas futures market.

Anatolyev, Seleznev, and Selezneva (2018) utilize the unique sequential nature of the API and DOE reports announcements in characterizing the formation of market participants' beliefs around these announcements. The authors conclude that the expectations are formed uniformly across the futures term structure curve and that there exists a strong and negative relationship between the announcement surprises and crude oil futures returns.

Based on the studies discussed above, we state that, due to the presence of periodic and significant announcements, the WTI futures electronic market is an interesting setting in which to study the dynamics of the VPIN measure and its potential to capture order flow toxicity. We next describe the WTI futures trade data and the oil inventory announcements data we use in this study.

2.3 Data

2.3.1 WTI Crude Oil Futures

Our data contains information on all trades in the WTI Light Sweet Crude Oil futures contracts (CME Globex and Bloomberg product code: CL) traded at the New York Mercantile Exchange (NYMEX), a commodity futures exchange owned and operated by CME Group (Chicago Mercantile Exchange). The NYMEX electronic order book futures market hours start at 6 p.m. Eastern Time (ET) and end at 5 p.m. ET, Sunday to Friday. The minimum price fluctuation is 1 cent per barrel, while one unit of the CL futures contract represents 1,000 US barrels. [7] Our data sample covers a period of three years, from January 3, 2017 to December 31, 2019, yielding 755 trading days.

NYMEX offers monthly futures contracts with expirations listed for the current year, the next 8 calendar years, and 2 additional consecutive contract months. Therefore, on any given day, there are at least 99 crude oil futures contracts with different expiration dates available for trading. For each day, we select the futures contract with the greatest volume traded on that particular day for our analysis. Figure 2.1 displays the time series of daily closing prices (in US

[7]The full technical specification can be found at the CME's website: `https://www.cmegroup.com/trading/energy/crude-oil/light-sweet-crude_contract_specifications.html`

dollars per barrel) and the daily total traded volume (in contract units, or "lots") of the most active (in terms of the total traded volume) WTI futures contract. We note that the average daily volume in our sample is 660,431 contracts. However, there is a considerable variation of the total daily trade activity present in our sample, as the daily standard deviation of 158,038 contracts documents.

For each trade entry, we possess information about the trade time (stamped to the nearest millisecond), price, volume, and aggressor side identification (depending on whether the trade was buyer- or seller- initiated). We note that the knowledge of the side of the order book on which each trade occurred constitutes valuable information that eliminates the neccessity to estimate the aggressor side of trade using classification algorithms (i.e., Lee and Ready (1991)).

Figure 2.2 displays the average daily traded volume profile at one-minute aggregation. The time interval displayed in Figure 2.2 is 3 a.m. - 5 p.m. ET (9:00 - 23:00 Central European Time (CET)), because this accounts for the absolute majority of the daily traded volume. We note that, not only does the number of contracts traded vary across days, but there is also significant variation in trading activity within each trading day. The structure of this intraday liquidity profile shows a rapid increase in trading activity occuring exactly at 9 a.m. ET, and a considerable surge just before its rapid decrease after 2:30 p.m. (corresponding to the crude oil pit opening and closing times at NYMEX, occurring at 9 a.m. and 2:30 p.m. ET). On average, at the peak time just before the pit closes, more than 4,000 units of the most liquid maturity contract are traded per minute.

2.3.2 US Crude Oil Inventories

Another type of data used in this study is the weekly announcement data on the levels of crude oil inventories in the United States. The two main announcements used by investors to assess changes in the amounts of inventories (and hence, weekly changes in the nation's overall demand for and the supply of crude oil) are the Weekly Statistical Bulletin report by the American Petroleum Institute (API) and the US Department of Energy (DOE) report by the Energy Information Administration (EIA). We refer to the two reports as the API and the DOE.

The usual schedule of report releases is as follows: the API report is released each Tuesday at 4:30 p.m. ET (usually 22:30 CET) and the DOE report is released each Wednesday at 10:30 a.m. ET (usually 16:30 CET). If a Monday is a holiday, the API report is released on the following Wednesday at 4:30 p.m. ET and the following DOE report is released on Thursday at 11:00 a.m. ET.

The industry group API represents various American manufacturers, producers, distributors, and refineries of crude oil and petroleum products. The Weekly Statistical Bulletin published by

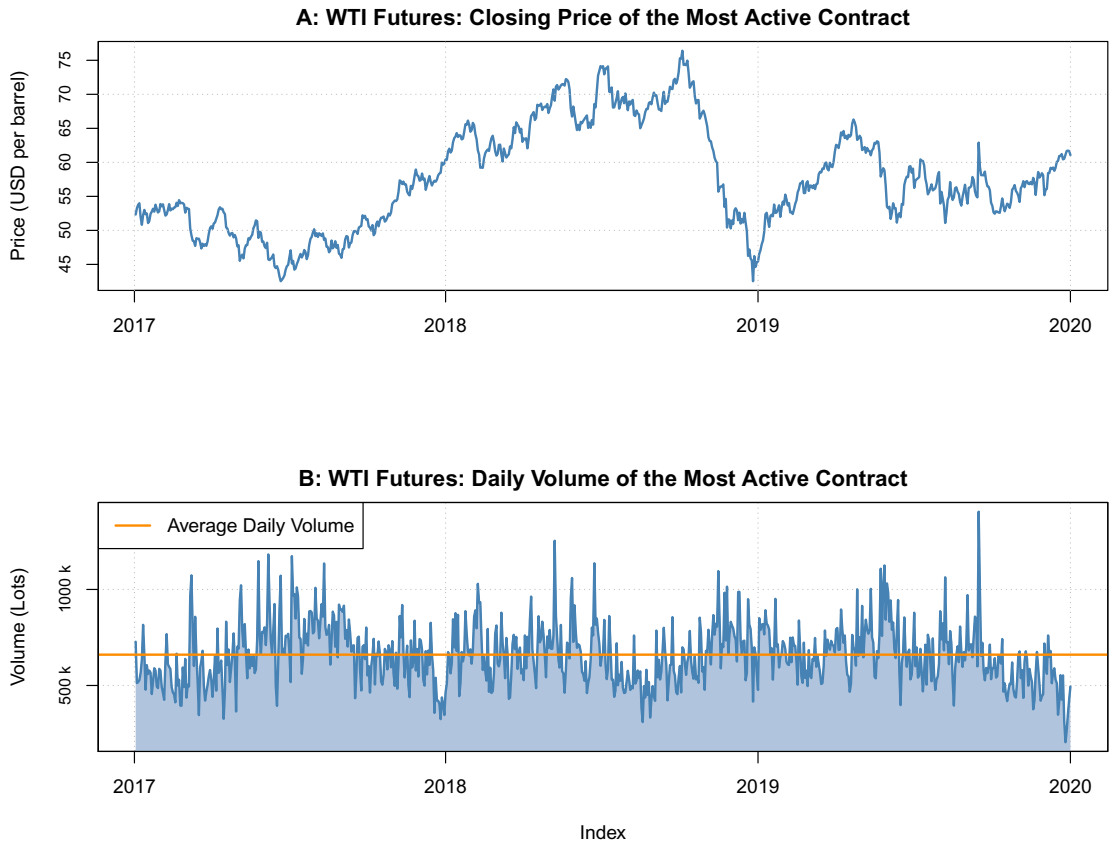


Figure 2.1: WTI futures most active contract time series. Panel A displays the daily closing prices in USD per US barrel. Panel B displays the daily traded volume of the most active futures contract. One unit (one lot) of the WTI futures contract represents 1,000 US barrels. Time period observed: January 3, 2017 - December 31, 2019. The horizontal line displays the sample's average daily volume of 660,431 lots.

this group is based on information that is disclosed by these companies largely on a voluntary basis. One of the major variables included in the report is the current overall level of crude oil inventories. The other related categories include, among others, motor gasoline, kerosene jet fuel, and distillate fuel oil inventories as well as data on imports, refinery input and container capacity. At the time of its publication, the data is available solely to subscribers through Thomson Reuters, however, historical data on the crude oil inventories changes reported in these documents is readily available online, at, for example, `https://www.tradingfloor.com` or `https://www.investing.com`.

Conversely, the Energy Information Administration publishes its weekly Petroleum Status Report at the institution's website and the key figures are immediately published via Bloomberg, Thomson Reuters, and other information agencies. Contrary to the voluntary information submission of the API report data, EIA directly requires major oil and petroleum related companies

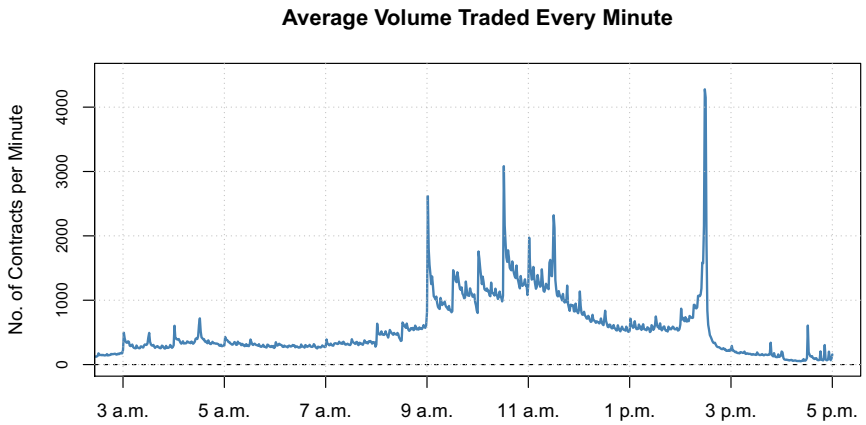


Figure 2.2: The average traded volume of the most active WTI futures contract by minute for the 3 a.m. - 5 p.m. ET (usually 9:00 - 23:00 CET) part of the trading day. This average daily profile is calculated throughout the January 3, 2017 - December 31, 2019 sample.

to submit their inventory levels information. Any intentional misreporting or non-compliance with these requirements is punishable by law. Beyond changes in total crude oil inventories, other published data include, for example, gasoline and distillate fuel inventories and total US crude oil refinery inputs, gasoline production, and crude oil imports.

We focus on the principal indicator published in both reports, the change in the total US crude oil inventories (expressed in barrels), excluding the inventories in the Strategic Petroleum Reserves. We collect the weekly data for the period from 2017 to 2019 for both the API report (we obtained the data from `https://www.investing.com/economic-calendar/api-weekly-crude-stock-656`) and the DOE report (we obtained the data from Bloomberg). The three year data sample used in this study contains 156 weekly API announcement values and 156 corresponding weekly DOE values. We also possess information on the actual absolute level of US crude oil inventories at the beginning of our sample - on January 5, 2017, a decrease in the US commercial inventories of 7.05 million barrels from the previous week was reported, and the absolute inventories amounted to 479.012 million barrels.[8]

A graphical and statistical summary of crude oil inventory data appears in Figure 2.3 and Table 2.1. Panel A in Figure 2.3 displays the overall level of crude oil inventories in the United States. Throughout our sample period, inventory levels reached a peak of 535.5 million barrels on April 5, 2017 and a trough of 394.1 million barrels on September 19, 2018. The remaining panels summarize the time series of the announced weekly changes in US crude oil inventories for the two reports, and the time series of their cross-differences. Not surprisingly, we observe that the API and DOE time series exhibit a substantial degree of positive cross-correlation (correlation coefficient equal to 0.74). However, weeks during which noticeable discrepancies between the

[8] `https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=WCESTUS1&f=W`

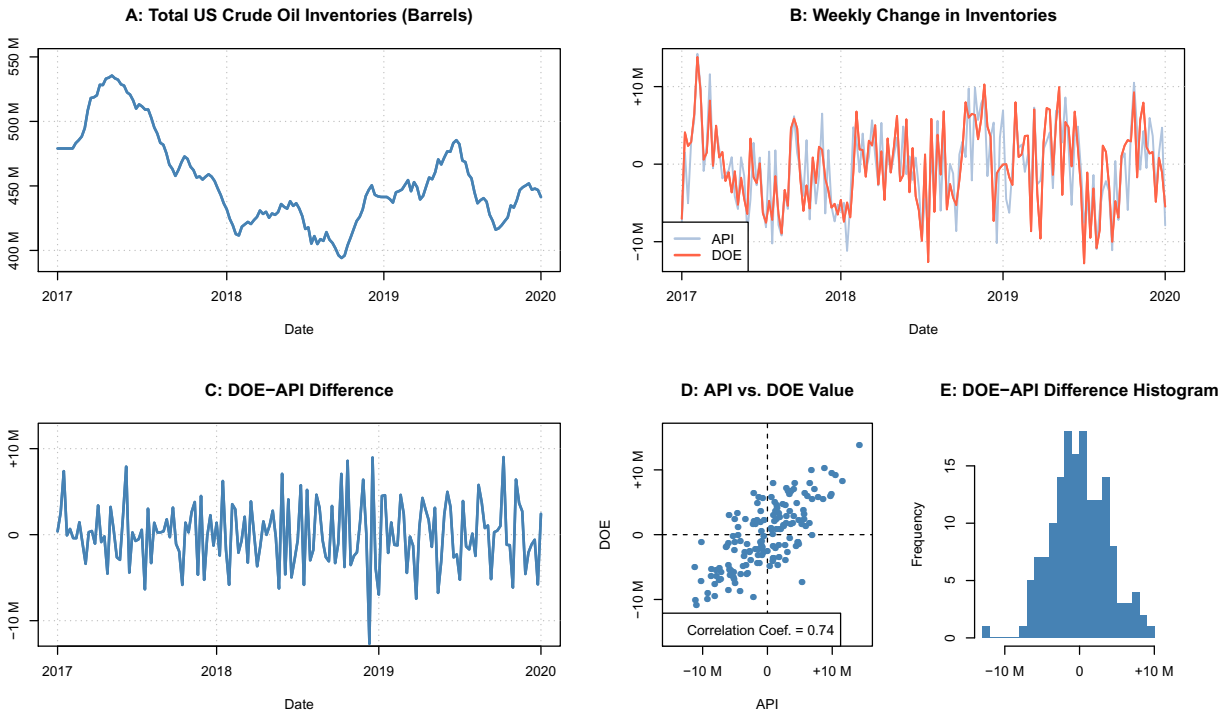


Figure 2.3: Upper panels: weekly time series of the level of US crude oil inventories and the changes in oil inventories as reported by the API and DOE reports. Lower panels: Time series of the weekly differences between the two reports and their corresponding scatter plot and histogram.

two reports exist are not uncommon. The largest difference in our sample corresponds to DOE data published on December 6, 2018, when the data reported a weekly decrease of 7.323 million barrels of inventories, while the API data published earlier the same week projected an increase of 5.360 million barrels, resulting in a discrepancy of more than 12 million barrels. The futures price reacted to this news with a sharp rise of more than 60 cents per barrel within minutes, equivalent to more than a 1 percent increase in price.

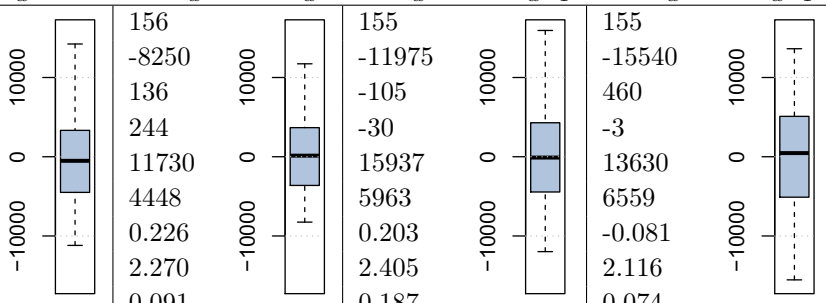
Denoting by API_w^A (DOE_w^A) the change in inventory levels as announced in the API (DOE) report during week w, we infer from Table 2.1 that, while both the API (column API_w^A) and the DOE (column DOE_w^A) time series exhibit relatively strong significant autocorrelation (of up to the third order), their cross-differences (column $DOE_w^A - API_w^A$) are negatively first-order autocorrelated. Moreover, the first differences of both inventory changes time series (columns $API_w^A - API_{w-1}^A$ and $DOE_w^A - DOE_{w-1}^A$) exhibit strong negative first-order autocorrelation, suggesting that a positive change in inventories is more likely to be followed by a negative change over the next week and vice versa. The overall distributions of the API and the DOE data are reasonably similar, spanning the interval of -14.5 to +14.5 million barrels, being symmetric roughly around zero, slightly positively skewed and having the standard deviation of approximately 5 million barrels.

Table 2.1: Summary statistics of API, DOE and the surprise candidates time series.

Summary statistics of the time series based on the reported and forecasted API and DOE inventory changes, and boxplots of the distributions of these time series. For week w, the actual API (DOE) value is labeled API_w^A (DOE_w^A) and the corresponding week's analysts' median forecasts of the DOE value is labeled DOE_w^F. The Min, Median, Mean, Max, and Standard Deviation statistics, as well as the displayed boxplot data, are expressed in thousands of US barrels. For the ACF, the (*) sign highlights values that are significant at the 5% level. JB represents the Jarque-Bera test.

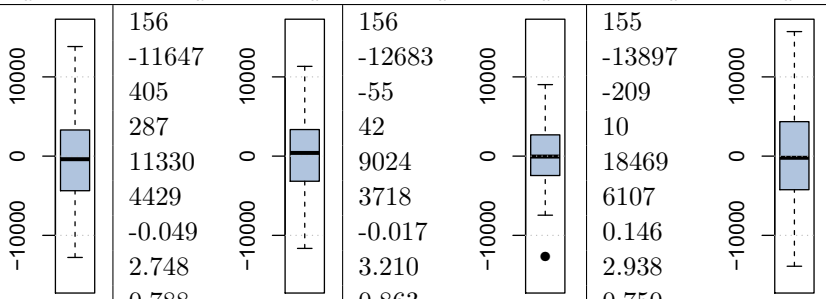
Panel A: API Report Time Series & Three Surprise Candidates

Time series:	API_w^A	$API_w^A - DOE_w^F$	$API_w^A - DOE_{w-1}^A$	$API_w^A - API_{w-1}^A$
No. of Obs.	156	156	155	155
Min	-11190	-8250	-11975	-15540
Median	-515	136	-105	460
Mean	-328	244	-30	-3
Max	14230	11730	15937	13630
St. Dev.	5195	4448	5963	6559
Skewness	0.082	0.226	0.203	-0.081
Kurtosis	2.578	2.270	2.405	2.116
JB p-value	0.513	0.091	0.187	0.074
ACF(1)	0.195*	-0.145	-0.450*	-0.518*
ACF(2)	0.248*	-0.012	-0.019	0.009
ACF(3)	0.270*	0.126	-0.026	0.111



Panel B: DOE Report Time Series & Three Surprise Candidates

Time series:	DOE_w^A	$DOE_w^A - DOE_w^F$	$DOE_w^A - API_w^A$	$DOE_w^A - DOE_{w-1}^A$
No. of Obs.	156	156	156	155
Min	-12788	-11647	-12683	-13897
Median	-390	405	-55	-209
Mean	-287	287	42	10
Max	13830	11330	9024	18469
St. Dev.	5213	4429	3718	6107
Skewness	0.009	-0.049	-0.017	0.146
Kurtosis	2.499	2.748	3.210	2.938
JB p-value	0.441	0.788	0.863	0.750
ACF(1)	0.310*	-0.048	-0.255*	-0.513*
ACF(2)	0.342*	0.116	-0.110	0.136
ACF(3)	0.188*	0.042	-0.054	-0.087



2.4 Methodology

2.4.1 Defining the Announcement Surprises

The level of crude oil inventories is one of the most closely watched oil market indicators. As is usually the case with publications of crucial macroeconomic data, these announcements are also accompanied by ex ante estimates of upcoming values. These estimates are usually based on surveys of professional analysts, forecasters, and market practitioners, and are released by various

information agencies during the days leading up to announcements of the official numbers. In our study, we also possess weekly survey median estimates for the DOE report data published by Bloomberg. These survey estimates are often used in the literature as prior market expectations of the announced values and the market reaction to the actual announcement is often evaluated with respect to market surprise, defined as the difference between the actual and the expected values of the announcement.[9]

Before we evaluate the impact of each announcement event, we must, however, specify each announcement's prior market expectation. Once the prior expectation is defined, the market surprise associated with each announcement is then simply calculated as the difference between the announced value and the prior expectation. In this section, we select the particular definitions of market prior expectations out of four possible candidates for each type of the report.

We possess three report-related time series: weekly announced API values (denoted API_w^A, where $w = 1, \ldots, 156$ indexes all weekly values in our three-year sample), as well as the weekly DOE values, also accompanied by DOE survey forecasts (denoted DOE_w^A and DOE_w^F). For each w, we also denote the set of all values known before the week's w API (DOE) announcement by I_w^{API} $\left(I_w^{DOE}\right)$. Assuming this notation, we consider the following four prior expectations to be candidates for both reports:

$$E\left[API_w^A | I_w^{API}\right] = \begin{cases} 0 \\ DOE_w^F \\ DOE_{w-1}^A \\ API_{w-1}^A \end{cases}, \qquad E\left[DOE_w^A | I_w^{DOE}\right] = \begin{cases} 0 \\ DOE_w^F \\ API_w^A \\ DOE_{w-1}^A \end{cases} \tag{2.1}$$

Therefore, for each report, we consider a naive market expecation of zero, an expectation equal to the reported median survey forecast of that week's DOE report, an expectation equal to the nearest previously announced inventory value (considering both report types pooled together) and an expectation equal to the previous reported value, considering each report separately. We note that, due to the commercial subscription nature of the API report, we do not possess survey data for the API announcements. However, since the corresponding DOE data is usually published within 24 hours following the API report, the DOE survey forecasts are always readily available before the API report is published. Therefore, the value DOE_w^F may be considered an element belonging to I_w^{API} and the second possible candidate for the API market expectation variable can be defined.

When a market expectation is specified, the market surprise associated with the week's w

[9]This approach is taken, for example, by Andersen et al. (2003), Bu (2014), Armstrong, Cardella, and Sabah (2017), and Bjursell, Wang, and Zheng (2017).

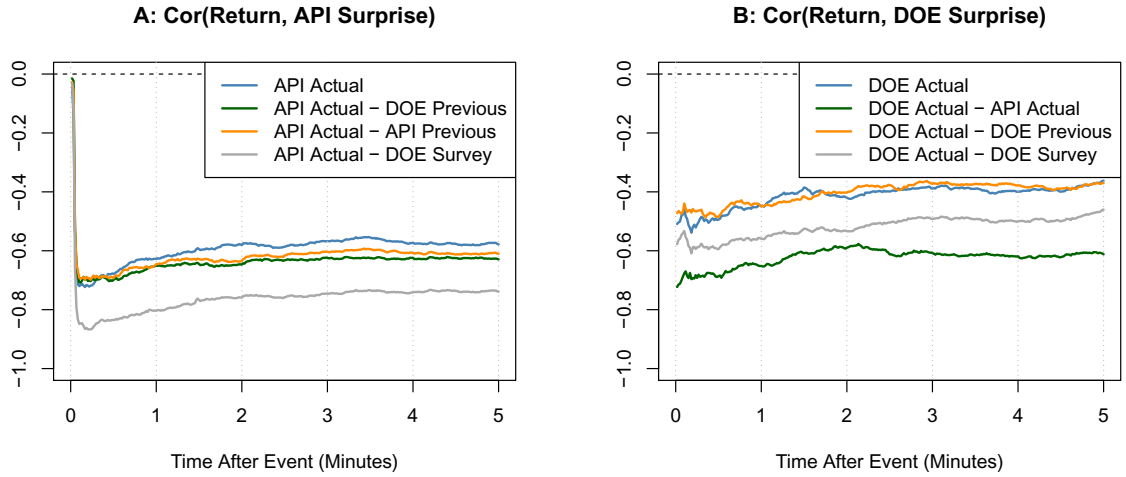


Figure 2.4: Correlation coefficients between the values $p_{t+s} - p_t$ and various (non-normalized) announcement surprise definitions for the range of values $s = 1, \ldots, 300$ and for both types of reports (Panel A: API, Panel B: DOE), where p_t is the log price of the WTI futures contract just before the announcement release and p_{t+s} is the log price s seconds after. For each panel, we consider four different surprise definitions as described in equation 2.1 and labeled in the legend. Each correlation coefficient is calculated based on 156 weekly (surprise value, $p_{t+s} - p_t$) pairs. We chose the maximum time horizon of 5 minutes for graphical illustration purposes, as the correlation trajectories stabilize beyond this time horizon.

API announced value is defined as:

$$SUR_w^{API} = API_w^A - E[API_w^A \,|I_w^{API}]\,, \tag{2.2}$$

and analogously for the DOE report. For the sake of robustness, we consider not only the surprise defined as the absolute difference between the announced and the expected value, but also its relative difference relative to the most recent total size of US inventories (denoted as INV_w):

$$SURN_w^{API} = \frac{API_w^A - E[API_w^A \,|I_w^{API}]}{INV_w}, \tag{2.3}$$

and, again, similarly for the DOE report.

The statistical properties of all four (non-normalized) surprise candidate time series for both reports are displayed in Table 2.1. We note that all time series are roughly centered around zero and overall possess similar distributional properties. While the original API and DOE reported values are significantly positively autocorrelated (at least up to the third lag), their first differences ($API_w^A - API_{w-1}^A$ and $DOE_w^A - DOE_{w-1}^A$, respectively) are strongly negatively autocorrelated within the first lag. The other surprise time series are either not autocorrelated at all, or are negatively first-order autocorrelated.

Out of the four surprise candidates for each report (defined in 2.1), we select the one that

correlates most closely with the post-announcement price discovery dynamics. Figure 2.4 displays the correlation coefficients between each of the four considered surprise candidates and the post-announcement log-returns of the WTI trade prices. The log-returns are calculated relative to the price prevailing immediately before the announcement, and are evaluated at a sequence of WTI prices sampled from one second to 300 seconds (5 minutes) after each announcement. The maximum time horizon of 5 minutes is chosen for graphical illustration purposes, as the correlation trajectories stabilize beyond this time horizon. Panel A considers the API announcement and the four non-normalized API surprise candidates. We note that all the estimated correlation coefficients are strongly negative, in line with our expectations, implying that a positive surprise (higher-than-expected inventory levels) tends to lead to a negative price shock. Not only does this price shock follow the announcement almost immediately, the price effect is persistent over the 5 minute period after the announcement as well. Out of the four surprise candidates, the one that is most strongly negatively correlated with the WTI price returns is the second surprise candidate, defined as $API_w^A - DOE_w^F$ (grey line). We conjecture that the second market prior candidate, the DOE survey forecast (DOE_w^F), is the most suitable choice for the market's prior expectations for the API report.

Similarly, Panel B of Figure 2.4 displays the surprise candidate - WTI return correlations for non-normalized surprise candidates for the DOE announcement. In this case, the third surprise candidate (green line) appears to be the most strongly negatively correlated in a persistent manner, with all the estimated correlations below -0.6 during the five minute interval after the announcement. We therefore choose the third surprise candidate, defined as $DOE_w^A - API_w^A$. When we consider the normalized surprise values, all the estimated correlations are almost identical, leading us to the same conclusions.

Interestingly, rather than the median value of the survey of professional forecasters (DOE_w^F), the same-week actual reported API value (API_w^A) is our choice for identifying the market prior expectation for the corresponding DOE report. We note that this definition of the announcement surprise is in line with the fact that the nature and the informational content of both reports should be similar, according to their respective data providers (as the API website states, "API collects an exact copy of the data submitted to EIA. Respondents send data to API using the same weekly survey forms that EIA uses. (...) It is true that companies voluntarily send API a copy of the data they send to EIA; however, the fact that they do so voluntarily is irrelevant. Both API and EIA are on record stating that their reported weekly data cover roughly 90% of the industry. Since API and EIA need only estimate the remaining 10%, differences in their weekly estimates can be largely attributed to statistical noise.").[10].

Thus, in this study, we employ the following definitions of the (normalized) surprises associ-

[10]See the Q&A at `https://www.api.org/products-and-services/statistics/api-weekly-statistical-bulletin`

ated with the weekly API and DOE report announcements:

$$SURN_w^{API} = \frac{API_w^A - DOE_w^F}{INV_w}, \quad SURN_w^{DOE} = \frac{DOE_w^A - API_w^A}{INV_w}. \tag{2.4}$$

2.4.2 The Classical PIN Model

In their seminal work on market microstructure, Easley et al. (1996) develop the following model: trading is viewed as a periodically repeated game between market makers (liquidity providers) that quote bid and offer prices, and market takers (aggressive traders), that react to these bid and offer quotes. Each market maker is assumed to be risk neutral and behaves competitively. At any point in time, the market maker quotes bid and offer prices, ready to buy or sell one unit of a risky asset.

We index the periods (for example, days) by $i = 1, \ldots, I$ and time within each period by $t \in [0, T]$. At the beginning of each period, two states of the world may occur: either an information event has occurred (with probability $\alpha \in (0, 1)$) or a no-information event has occurred (with probability $(1 - \alpha)$). These events are generated independently across periods. Assuming that in period i a "good news" information event has occurred, the underlying asset is assumed to be worth $\bar{V}_i$ and the probability of the "good news" event equals $(1 - \delta)$. Correspondingly, the probability of a "bad news" event, under which the asset is worth $\underline{V}_i < \bar{V}_i$, is assumed to be equal to $\delta \in (0, 1)$. After the end of each period, the value of the asset is realized and made public. If no news event has been realized, the prior value of the asset is assumed to be $V_i^* = \delta \underline{V}_i + (1 - \delta)\bar{V}_i$, satisfying $\underline{V}_i < V_i^* < \bar{V}_i$.

There are two types of aggressive traders: informed and uninformed. Given the three possible states of the world (good news, bad news, no news), trade orders from informed investors are assumed to arrive according to the Poisson process at rate μ. These informed investors always buy under the good news event, always sell under the bad news event, and do not trade under the no news event. Uninformed buy and sell traders arrive under all three possible states of the world according to the Poisson process at rate ε. All Poisson processes are mutually independent.

The market maker knows the values of $\alpha, \delta, \mu, \varepsilon$ and therefore also knows the ex ante information events probabilities (equal to $1-\alpha, \alpha(1-\delta)$,and $\alpha\delta$ for the good news, bad news, and no news events, respectively). However, he does not know which state of the world has occurred. The market maker therefore uses Bayesian updating in order to infer the ex post probabilities of information events based on the types (buy or sell) and rates of arrival of the aggressive traders' orders. We denote the market maker's prior belief about the no news, good news, and bad news state of the world at time t as $P(t) = (P_n(t), P_g(t), P_b(t))$.

At time t, the market maker sets the bid and ask prices (B_t, A_t) in such a way that these prices correspond to the expected value of the asset conditional on the market maker's time-t

prior beliefs $P(t)$ and the fact that a counterparty is willing to sell or buy. For example, assuming that a sell order arrives at time t (and denoting this event as S_t), by the Bayes rule, the market maker's updated (posterior) belief about the no news event at time t equals:

$$P_n(t|S_t) = \frac{P_n(t)\varepsilon}{\varepsilon + P_b(t)\mu}.$$

The corresponding conditional probabilities of good and bad news events may then be expressed as:

$$P_g(t|S_t) = \frac{P_g(t)\varepsilon}{\varepsilon + P_b(t)\mu}, P_b(t|S_t) = \frac{P_b(t)(\varepsilon + \mu)}{\varepsilon + P_b(t)\mu}.$$

The market maker's bid price on day i at time t is thus equal to:

$$B_t = P_n(t|S_t)V_i^* + P_g(t|S_t)\bar{V}_i + P_b(t|S_t)\underline{V}_i = \frac{P_n(t)\varepsilon V_i^* + P_b(t)(\varepsilon + \mu)\underline{V}_i + P_g(t)\varepsilon\bar{V}_i}{\varepsilon + P_b(t)\mu}.$$

Analogously, the ask price may be written as:

$$A_t = \frac{P_n(t)\varepsilon V_i^* + P_b(t)\varepsilon\underline{V}_i + P_g(t)(\varepsilon + \mu)\bar{V}_i}{\varepsilon + P_g(t)\mu}.$$

Algebraic manipulation of these expressions leads to the following bid-ask spread at time t, denoted as $\Sigma_t = A_t - B_t$:

$$\Sigma_t = \frac{P_g(t)\mu}{\varepsilon + P_g(t)\mu}\left(\bar{V}_i - E[V_i|t]\right) + \frac{P_b(t)\mu}{\varepsilon + P_b(t)\mu}\left(E[V_i|t] - \underline{V}_i\right), \tag{2.5}$$

with the conditional expected value of the asset at time t being equal to:

$$E[V_i|t] = P_n(t)V_i^* + P_g(t)\bar{V}_i + P_b(t)\underline{V}_i.$$

Equation 2.5 implies that the bid-ask spread is equal to the sum of the probability of an informed buy trade, multiplied by the expected loss incurred due to the fact that the buyer is informed, and of the symmetric term corresponding to an informed seller. The market maker thus sets the spread in order to protect against the adverse selection he is exposed to because of the possibility of the arrival of an informed trader. This probability of an informed trade (PI) is the central element of the model and may be written as the sum of the probabilities of informed buy and sell trades expressed in equation 2.5, that is:

$$PI(t) = \frac{P_g(t)\mu}{\varepsilon + P_g(t)\mu} + \frac{P_b(t)\mu}{\varepsilon + P_b(t)\mu} = \frac{\mu(1 - P_n(t))}{\mu(1 - P_n(t)) + 2\varepsilon}.$$

Assuming that the probabilities of good and bad events are equal ($\delta = 0.5$), the probability

of informed trading at time $t = 0$ equals:

$$PI(0) = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \tag{2.6}$$

and the spread at the day open becomes:

$$\Sigma_0 = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \left[\bar{V}_i - \underline{V}_i\right].$$

In this model, Easley et al. assume that we observe the total numbers of buyer and seller initiated trades for each day: $\{X_i, Y_i\}_{i=1}^{I}$. Based on this observed data, the parameter vector $M = (\alpha, \delta, \mu, \varepsilon)$ can be estimated by maximizing the product of the daily likelihood functions:

$$L\left(M|\{X_i, Y_i\}_{i=1}^{I}\right) = \prod_{i=1}^{I} L\left(M|X_i, Y_i\right),$$

where the daily likelihood function can be written as a weighted sum of observing X buys and Y sells on the no news, bad news, and good news days, with the weights corresponding to the probabilities of such days occurring:

$$\begin{aligned} L\left(M|X,Y\right) = (1-\alpha)e^{-\varepsilon T}\frac{(\varepsilon T)^X}{X!}e^{-\varepsilon T}\frac{(\varepsilon T)^Y}{Y!} + \alpha\delta e^{-\varepsilon T}\frac{(\varepsilon T)^X}{X!}e^{-(\mu+\varepsilon)T}\frac{((\mu+\varepsilon)T)^Y}{Y!} \\ + \alpha(1-\delta)e^{-(\mu+\varepsilon)T}\frac{((\mu+\varepsilon)T)^X}{X!}e^{-\varepsilon T}\frac{(\varepsilon T)^Y}{Y!}. \end{aligned}$$

2.4.3 The VPIN Model

The PIN metric assumes that the beginning of day probability of informed trading (denoted as $PI(0)$, equation 2.6 above) is constant over the observed period $i = 1, \ldots, I$. Therefore, assuming that the probabilities of information events and the corresponding rates of order flow are constant as well, such "stationarity" of the model can be used to directly estimate $PI(0)$.

Easley, de Prado, and O'Hara (2010) enhance the PIN metric in a way that enables more frequent updating of the probability of informed trading throughout the day. This approach accounts for the fact that information events do not arrive regularly, can arrive multiple times each day, and can differ in their importance and potential to affect the current value of an asset. In today's futures markets, which are dominated by high-frequency trading, such aspects of information arrival need to be accounted for in any metric that aims to estimate the probability of informed trading and the degree of adverse selection that it accompanies.

The main result that serves as a transition to the new metric is the analysis by Easley et al. (2008), in which informed and uninformed traders' arrival rates are no longer assumed to be constant, but rather follow a time-varying probability distribution, similar to a GARCH-type

process. Under these assumptions, the authors show that, for a given time period, the PIN can be approximated as follows:

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \approx \frac{E\left[V_S - V_B\right]}{E\left[V_S + V_B\right]}. \tag{2.7}$$

where V_B, V_S represent the aggregate volume of buyer and seller initiated trades over the period. Informed trade intensity $\alpha\mu$ from equation 2.6 is therefore approximated by the expected volume imbalance $E\left[V_S - V_B\right]$ in equation 2.7, while the total trade intensity $\alpha\mu + 2\varepsilon$ is approximated by the expected total volume $E\left[V_S + V_B\right]$.

The second important building block in developing the enhanced metric is the transfer from clock-time to volume-time. In particular, instead of assuming that the probability of informed trading remains constant over time, the new metric allows for this probability to be updated each time a pre-determined amount of volume is traded. Easley, de Prado, and O'Hara (2010) argue that a given amount of traded volume (also referred to as a volume bucket) represents a given amount of corresponding new information. When there is relatively little new information, trade (and volume) intensity is smaller, and it takes a longer period of time before an entire volume bucket is traded. Likewise, when new information arrives, informed traders have an obvious incentive to trade on this information as quickly as possible, and trade intensity increases. Consequently, additional volume buckets are traded in a shorter period of time. Arguably, the more relevant and more surprising the new information is, the greater the increase in trading intensity will be, and the volume buckets fill more quickly. The new metric allows for updates of the estimated probability of informed trading after each volume bucket completion, which, based on the choice of the volume bucket size, can occur many times per day. Therefore, this enhanced metric is referred to as the Volume-synchronized Probability of Informed Trading (VPIN).

The VPIN calculation procedure can be summarized as follows:

1. Inputs: The information about trades containing the time, volume, and the aggressor side (buyer or seller initiated) of each trade.[11]

   Parameters: the size of one volume bucket (B) and the length of the moving average required to calculate one VPIN measure (n).

2. The trade data is aggregated into consecutive volume buckets, where each volume bucket contains an equal number of contracts traded (chosen as B). If the volume of a trade that would complete a volume bucket is greater than the volume remaining in that bucket, the trade is divided into appropriate sub-trades, and each sub-trade is treated individually. On any given trading day, this procedure yields, say, N volume buckets $b_1, \ldots, b_N$. The

[11] When the information on the aggressor side of the trades is not available, a trade classification algorithm, such as the Lee and Ready (1991), or the Easley, de Prado, and O'Hara (2012) bulk classification algorithm can be used.

sum of traded volume of each bucket is equal to B. The remaining trades left at the end of the day that would not fill an entire volume bucket are disregarded.

3. For each volume bucket b_i, we calculate the total volume of buyer initiated trades and denote it as $V_B^{(i)}$. Similarly, the total volume of seller initiated trades will be $V_S^{(i)} = B - V_B^{(i)}$. The order imbalance corresponding to the volume bucket b_i is then defined as the absolute value of the difference of the total traded volume of buyer and seller initiated trades:

$$OI^{(i)} = |V_B^{(i)} - V_S^{(i)}|$$

4. The VPIN, known after volume bucket b_i has been completed, is then calculated as the moving average of the last n order imbalances:

$$VPIN_i = \frac{1}{n} \sum_{j=0}^{n-1} \frac{\left(|V_B^{(i-j)} - V_S^{(i-j)}|\right)}{\left(V_B^{(i-j)} + V_S^{(i-j)}\right)} = \frac{\sum_{j=0}^{n-1} OI^{(i-j)}}{nB}, \quad i \geq n.$$

We note that the VPIN is an asynchronous measure by definition because it is not updated regularly each predetermined number of seconds. Rather, a predetermined number of contracts need to be traded before another update of the VPIN value arises.

2.4.4 Hypotheses Formulation

We apply the VPIN calculation procedure described above to the WTI crude oil futures trade data at our disposal. A particular interest of this study is in quantifying the dynamics of the VPIN indicator around the US crude oil inventory announcement dates, taking into account the two kinds of inventory reports (DOE and API) separately. Publication of the crude oil inventory data is generally accepted as a major macroeconomic event with considerable market impact (Chang, Daouk, and Wang (2009), Bu (2014), Anatolyev, Seleznev, and Selezneva (2018)). This study also documents that rapid and substantial changes in the post-announcement price discovery process are indeed observable in the WTI futures market for both API and DOE reports. These regular announcement events can therefore function as an appropriate setting for event study investigation of the VPIN indicator dynamics. In this section, we formulate the general null hypotheses the validity of which we aim to investigate. We also provide their testable versions and descriptions of the empirical testing strategies we employ.

- **Hypothesis 1**

  On API (DOE) announcement days, the probability of informed trading as measured by VPIN before an API (DOE) announcement does not significantly differ from the value after the API (DOE) announcement.

Hypothesis 1 aims to test the essential required property of the VPIN metric, that it is able to capture the dynamics of informed order flow. Because we hypothesize that inventory data publications are accompanied by an increase in the intensity of informed order flow, we expect that VPIN will, on average, tend to increase in value shortly after the announcement. Moreover, we expect that the probability distribution of such VPIN changes will not be symmetric around zero.

In order to empirically test this hypothesis, we employ two paired tests: the paired Student's t-test and the non-parametric Wilcoxon signed rank test. We note that a similar approach has been used, for example, in event studies by Dimpfl (2011) and Bash and Alsaifi (2019). Considering one announcement type, for week w's announcement day, we calculate the overall level of VPIN before the announcement as the average of the last m VPIN values observed immediately before the announcement time, and denote this value as $VPIN_w^{m,Before}$. Similarly, we calculate the overall post-announcement VPIN as the average of the first m VPIN values observed immediately after the announcement and denote it as $VPIN_w^{m,After}$. For a fixed value of m, this procedure yields at most 156 pairs of $(VPIN_w^{m,Before}, VPIN_w^{m,After})$, indexed by $w = 1, \ldots, 156$, corresponding to the 156 weeks of sample data. For robustness purposes, we consider the following grid of values of the parameter $m : 1, 2, 3, 5, 10$.

For a given m and a given inventory report category, the Student's t-test version of the null hypothesis then becomes:

$$H_0^{(1a)} : E\left[VPIN_w^{m,Before} - VPIN_w^{m,After}\right] = 0. \tag{2.8}$$

Likewise, we use the non-parametric paired Wilcoxon signed rank test to test the following null hypothesis:

$$H_0^{(1b)} : \text{The distribution of } VPIN_w^{m,Before} - VPIN_w^{m,After} \text{ is symmetric around zero.} \tag{2.9}$$

The following Hypotheses 2 and 3 are evaluated separately for the API and for the DOE report using linear regression estimations. We use the Newey-West estimator of the covariance matrix of the coefficient estimates to account for possible heteroskedasticity and autocorrelation in the residuals. In this respect, we follow recent VPIN studies by Cheung, Chou, and Lei (2015), Armstrong, Cardella, and Sabah (2017) and Yildiz, Van Ness, and Van Ness (2020). We describe the definitions of the variables entering each regression below.

- **Hypothesis 2**

a) On API announcement days, the magnitude of the VPIN change around an API announcement does not depend on the magnitude of the API surprise.

b) On DOE announcement days, the magnitude of the VPIN change around a DOE announcement does not depend on the magnitude of the DOE surprise and does not depend on the magnitude of the corresponding previous API surprise.

We thus hypothesize that the more surprising the announced inventory value is relative to the previously defined market expectation, the greater the incentive of market participants to trade on the new information as quickly as possible will be, and hence the greater the magnitude of the post-announcement informed order flow increase will occur. Rejection of Hypothesis 2 will lead us to assert that the VPIN metric indeed captures the potentially rapid changes in the informed trading intensity in a continuous and directional manner. To empirically assess this hypothesis, we estimate the following set of linear regression models, regressing post-announcement changes in VPIN on the announcement surprises variables.

Considering first the API announcement, for each announcement day w and for a given m, we calculate the overall change in the magnitude of VPIN as the difference between the average of m VPIN values following immediately after the announcement, and the average of m VPIN values immediately preceding the announcement (this is the same as for Hypothesis 1) and define this difference as $Y_w^{(m,API)} = VPIN_w^{m,After} - VPIN_w^{m,Before}$. We relate this change in VPIN around the announcement to the normalized API surprise variable $SURN_w^{API}$, defined in 2.4. In order to establish the robustness of the results, we again iterate the parameter m over the grid of values 1, 2, 3, 5, 10. Since we hypothesize that both large positive and large negative values of surprises tend to be associated with greater increases in informed trading activity, we assume a quadratic relationship between the announcement surprise and the change in VPIN. Therefore, for a given m, we estimate the following linear regression:

$$Y_w^{(m,API)} = \alpha^{(m,API)} + \beta^{(m,API)} SURN_w^{API} + \gamma^{(m,API)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,API)} \quad (2.10)$$

Similarly, for the DOE report and a given m, we calculate the values $Y_w^{(m,DOE)}$ analogically to the API case and the magnitude of the DOE surprise $SURN_w^{DOE}$ defined in 2.4. Because we also aim to estimate the effects of the preceding API surprise on changes in VPIN around the DOE announcement, we estimate the following regression model:

$$\begin{aligned} Y_w^{(m,DOE)} = {} & \alpha^{(m,DOE)} + \beta^{(m,DOE)} SURN_w^{DOE} + \gamma^{(m,DOE)} \left(SURN_w^{DOE}\right)^2 \\ & + \delta^{(m,DOE)} SURN_w^{API} + \rho^{(m,DOE)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,DOE)} \end{aligned} \quad (2.11)$$

The testable versions of the null hypotheses then correspond to joint F-tests of significance of all the regression coefficients with the exception of the intercept in regression models 2.10 and 2.11:

$$H_0^{(2,API)} : \beta^{(m,API)} = \gamma^{(m,API)} = 0. \tag{2.12}$$

$$H_0^{(2,DOE)} : \beta^{(m,DOE)} = \gamma^{(m,DOE)} = \delta^{(m,DOE)} = \rho^{(m,DOE)} = 0. \tag{2.13}$$

- **Hypothesis 3**

  On API (DOE) announcement days, the change in the value of VPIN shortly before an API (DOE) announcement is not associated with the magnitude of the corresponding API (DOE) surprise.

  In this hypothesis, we are interested in the potential predictive ability of the changes in the intensity of informed trading shortly before the announcement time on the magnitude of the corresponding announcement surprise. First, we need to define a metric that will capture pre-announcement changes in VPIN behaviour. We do this by taking the difference of the last recorded pre-announcement VPIN value ($VPIN_w^{(Last,API)}$) and the moving average of the m VPIN values immediately preceding this last recorded pre-announcement value. We denote this moving average as $AVGVPIN_w^{(m),API}$. In this case, we iterate the parameter m through the values 1, 2, 3, 4, 5. Given m, the dependent variable for the API regression is then denoted as $X_w^{(m,API)} = VPIN_w^{(Last,API)} - AVGVPIN_w^{(m),API}$. The values for the DOE report are again calculated analogically.

  A relatively large value of X thus represents a relatively large increase in informed trading activity (as measured by VPIN) occurring shortly before the announcement is released. The dependent variables, representing the magnitudes of the surprises, are calculated as the absolute values of the normalized surprises defined in 2.4. Therefore, we do not distinguish between large positive and large negative surprises, because we assume that both types of surprises may be exploitable by potential informed traders. We next estimate the following regression models:

$$\begin{aligned} |SURN_w^{API}| &= \alpha^{(m,API)} + \beta^{(m,API)} X_w^{(m,API)} + \varepsilon_w^{(m,API)} \\ |SURN_w^{DOE}| &= \alpha^{(m,DOE)} + \beta^{(m,DOE)} X_w^{(m,DOE)} + \varepsilon_w^{(m,DOE)} \end{aligned} \tag{2.14}$$

  The testable versions of the null hypotheses thus correspond to the significance tests of the slope regression coefficients:

$$H_0^{(3,API)} : \beta^{(m,API)} = 0. \tag{2.15}$$

$$H_0^{(3,DOE)} : \beta^{(m,DOE)} = 0. \tag{2.16}$$

As noted earlier, we apply the Newey-West standard errors estimator to account for the potential heteroskedasticity and autocorrelation in the residuals whenever possible. Additionally, we note that, in all regression models, we also investigate the sensitivity of the results to the inclusion of additional explanatory variables, insofar as the relatively limited sample sizes of 156 weekly observations permit. In particular, we include quarterly dummy variables and the overall daily standard deviation of the WTI log-returns as additional regressors. However, we find that the results are highly robust to the inclusion of these variables and for the sake of brevity, we report the regression results corresponding to the equations specified above.

2.5 Results

We now analyze potential relationships between the weekly published levels of crude oil inventories for both reports and the WTI futures price discovery process, both from the price adjustment, as well as from the order flow toxicity (as measured by VPIN) perspectives.

2.5.1 Market Reaction to Announcements

In this section, we graphically summarize the price and volume reactions of the WTI futures contracts to both announcement events. To illustrate the strength and importance of these events as perceived by market participants, we also compare these characteristics with the corresponding time periods on days during which no announcements occurred.

Figure 2.5 displays the time development of the most liquid WTI futures contract price, sampled at one second frequency for the time window starting 30 minutes before the announcement and ending 30 minutes after the announcement for both API (Panel A) and DOE (Panel B) announcement events. For comparison purposes, price trajectories on all days without the two announcement events during the corresponding time windows are displayed in Panel C (non-API days) and Panel D (non-DOE days). Panels A and B display 156 time series each, while panels C and D display 599 time series each. Each price time series is normalized so that its value at the particular reference (event) time equals 1. Therefore, by design, all time series cross the value of 1 at this reference time.

Time series displayed on Panels A and B are colored according to the magnitude of the announcement event surprises, defined earlier as $API_w^A - DOE_w^F$ for the API report and $DOE_w^A - API_w^A$ for the DOE report. The positive values of the surprises (representing surplus inventories) are distinguished using blue shades, while the negative values (representing inventory shortages) are distinguished with green.

Additionally, Figure 2.6 displays the average traded volume by the number of WTI futures contracts per minute during the time window starting 30 minutes before and ending 30 minutes

after each announcement time, again for both types of inventory reports and for the corresponding time windows on non-announcement days. The lower panels show the average traded volume trajectory as accumulated from the time of the event to the future and to the past.

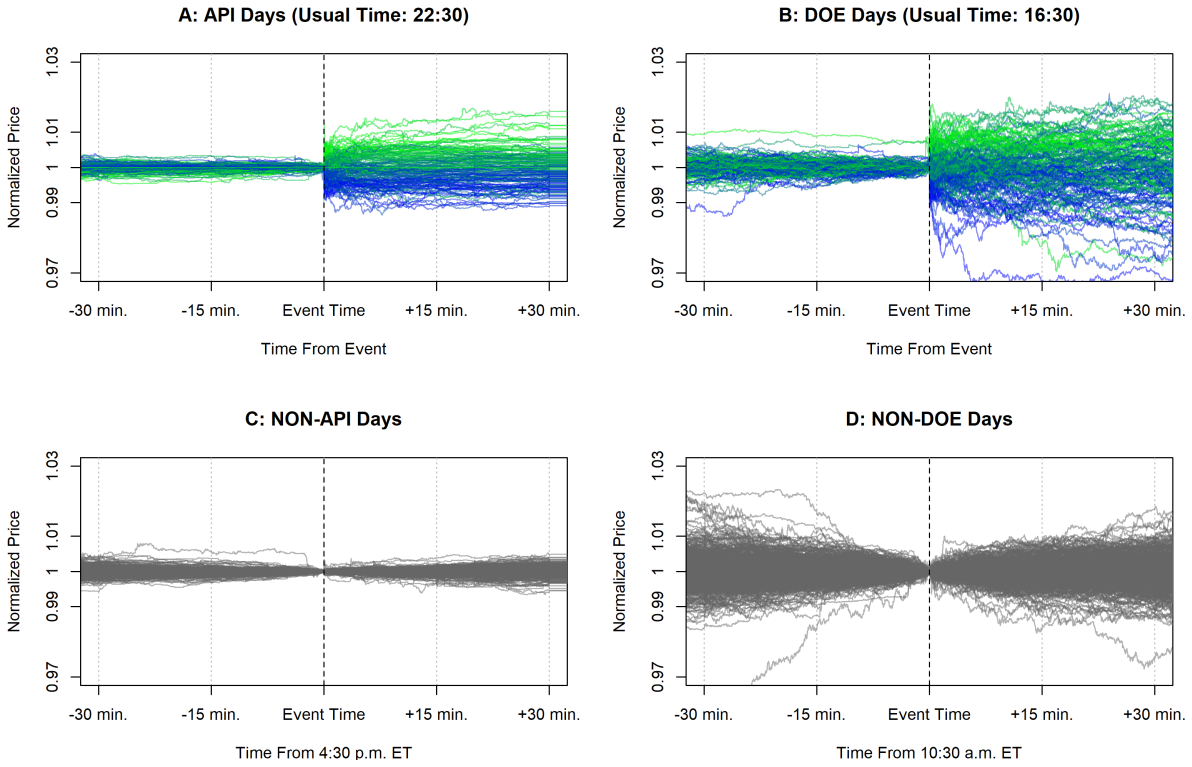


Figure 2.5: Normalized WTI price trajectories on API (Panel A) and DOE (Panel B) announcement days during time windows including 30 minutes before and 30 minutes after announcement times. The corresponding time windows on non-announcement days are displayed for comparison in Panel C and Panel D. Each time series is normalized so that the price at the time of the announcement equals 1. Therefore, all time series cross the value of 1 at Event Time. In the upper panels, the time series are colored according to the surprise value of the corresponding announcement, defined as $API_w^A - API_w^F$ for the API report and $DOE_w^A - API_w^A$ for the DOE report. The positive values of the surprises (representing surplus inventories) are designated by the blue end of the spectrum, while the negative values (representing inventory shortages) are designated by green.

It is apparent (and not surprising) from Figure 2.5 and Figure 2.6 that there exist systematic differences in WTI futures price discovery and trading activity between days with and without announcements. In particular, for both types of announcements, a sudden surge in price volatility (accompanied by a dramatic increase in traded volume immediately following the announcement) is observed at event time, while no such phenomena are observed for non-announcement days. The upper panels of Figure 2.5 suggest that the post-announcement price volatility increases are of the same order of magnitude for both the API and DOE reports, indicating that both reports are regarded as important gauges of the current state of the oil market.

Moreover, as the color spectrum of the announcement days time series suggests, the price

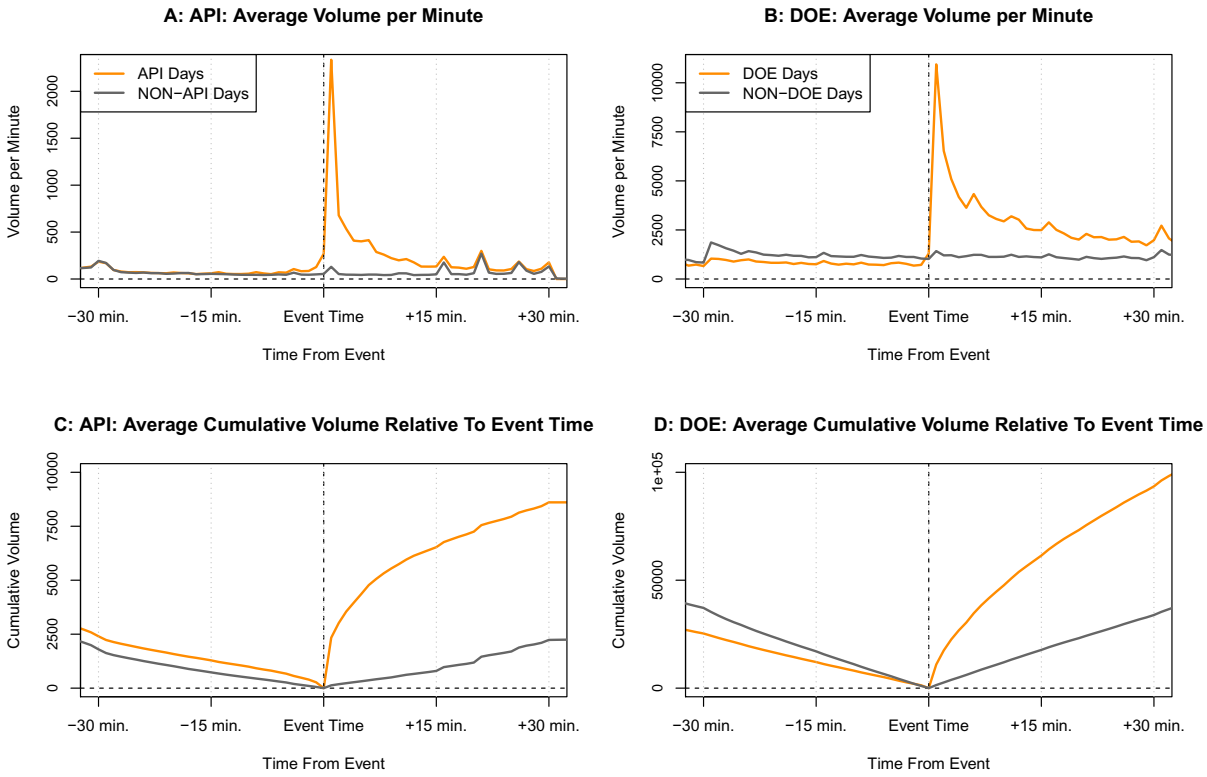


Figure 2.6: Upper panels: Average traded volume per each minute during the time window including 30 minutes before and after each announcement, and of the corresponding time windows on non-announcement days. Lower panels: Average cumulative traded volume during the same time windows. The event time is perceived as the origin, thus, the pre-event time period displays the average traded volume from the x-axis time until the event time. Left panels: API report. Right panels: DOE report.

reaction at these announcement times tends to be positive on days with negatively-valued announcement surprises and vice versa. This observation is also intuitively appealing. For example, a surprisingly small oil inventory level is associated with surprisingly high demand (or surprisingly low supply) of oil, leading to a price increase. We perceive this as an indication that our defined surprise measures do, in fact, capture the quality of new information that is being priced in the market after the announcements.

Although price volatility increases can be considered to be of the same magnitude, this relationship does not apply to the observed average traded volume trajectories. The approximately 10-fold post-announcement surge in average traded volume (as opposed to non-announcement days), though present for both reports, represents smaller absolute values of traded volume for the API than for the DOE report. This is again not surprising, since the DOE data is usually announced during market hours, at 10:30 a.m. ET and the API data is announced after the trading pit closes, during times of considerably lower market depth (the number of contracts quoted in the limit order book at any given price) and activity. Indeed, while on average, approximately 25,000 contracts are traded in the 30 minutes leading up to the DOE announcement,

this figure is 10 times smaller for the respective time window before the API announcement. We assert that the API's lower post-announcement traded volume is likely not observed because the report is less significant, but rather is due to lower market depth and greater bid-ask spreads that are usually present during the API announcement, requiring fewer contracts to be traded in order to cause price movements of a magnitude similar to that of the DOE report. [12]

Interestingly, we note that the price volatility during the DOE pre-announcement time window on DOE announcement days is considerably smaller than on non-announcement days. Additionally, as Panels B and D on Figure 2.6 show, the average pre-announcement traded volume on announcement days is approximately 30% smaller as well (about 26,000 contracts on DOE days vs. 40,000 contracts on non-DOE days between 10:00 a.m. and 10:30 a.m. ET). Following the DOE announcement, this relationship unsurprisingly reverts, leading to almost 100,000 contracts being traded during the following 30 minutes, while on average 35,000 contracts are traded on non-DOE days during the same period. We suggest that the scheduled DOE publication on announcement days alters the usual (non-announcement day) structure of market participant types engaging in trading during this time period. We hypothesize that market participants have greater incentives to lower their risk exposure to the oil market in the face of approaching DOE publication. Such behaviour would be associated with overall lower traded volume and a greater relative proportion of liquidity-demanding non-informed trades. In this sense, we are particularly interested in the information provided by the VPIN indicator.

Finally, we note that, while both the price and the volume change rapidly and immediately after the reports are published, no systematic change is observable during the minutes (or seconds) leading up to announcement times. Therefore, trading activity associated with the inventories information release appears to be present only after the information is made public, suggesting no presence of insider trading activity prior to the announcements. The main objective of the remainder of this paper is to investigate in greater detail whether the VPIN metric provides additional information about the structure and dynamics of information-based trading around the two inventory levels announcements.

2.5.2 VPIN

We apply the Volume-synchronized probability of informed trading (VPIN) metric to our WTI futures trade data spanning the period from January 2, 2017 to December 31, 2019. We reiterate that calculation of the VPIN metric requires a researcher to choose two input parameters: the size of one volume bucket B (the number of contracts to be traded before the VPIN metric is recalculated) and the number of past volume buckets n, used for the calculation of the moving average required to produce one VPIN value. There is no straightforward rule that would point

[12] We unfortunately do not possess the data on the structure and depth of the WTI limit order book.

to an optimal choice of the two parameters. Easley, de Prado, and O'Hara (2010) start with B corressponding to 2 percent of the average daily volume traded. This procedure therefore yields on average 50 volume buckets per day. The authors first set $n = 50$, but later adjust this parameter to different values.

We fix the value of $n = 30$, and for robustness purposes consider a grid of 5 different bucket sizes $B_i, i = 1, \ldots, 5$ corresponding to $0.1\%, 0.2\%, \ldots, 0.5\%$ of average daily traded volume.[13] These percentages of daily volume correspond to the following absolute sizes of buckets (in lots): 660, 1,321, 1,981, 2,642 and 3,302. For the sake of brevity, all remaining results are reported for the first bucket size category: $B_1 = 660$, corresponding to 0.1% of the average daily volume. Additionally, for robustness purposes, we report all results obtained in this chapter, applied to the third bucket size category $B_3 = 1981$ (corresponding to 0.3% of the average daily volume) in the chapter's Appendix 2.A.

We display the summary visual representation of the VPIN metric dynamics around inventory announcements in Figure 2.7. The first four panels (Panel A - Panel D) display the results expressed in the clock-time horizontal axis representation. The average VPIN trajectories for the API report (Panel A) and for the DOE report (Panel B) during a one hour window surrounding announcement times are shown, together with corresponding non-anouncement days trajectories. Additionally, the point standard deviation bands illustrate the degree of variability of these VPIN trajectories. Due to the high-frequency nature of the WTI futures market, Panels C and D show the zoomed-in one minute window around the announcement time. Panels E and F provide an alternative perspective by displaying the results represented in volume-time, showing the 50 average VPIN values obtained immediately before and immediately after each announcement.

Several observations can be made from Figure 2.7. Firstly, there clearly exist systematic differences in informed trading activity as measured by the VPIN metric between announcement and non-announcement days. While no systematic change in the behaviour of VPIN occurs on non-announcement days (as documented by the blue elements of Figure 2.7), both kinds of crude oil inventory reports announcements cause the VPIN to change considerably. The occurrence of the arrival of this new information causes, on average, an immmediate several percentage point increase in the estimated informed trading probability. Both reports appear to increase this probability by at least 3 percentage points, representing approximately a 20% increase. However, the average profiles of the informed trading activity exhibit different patterns for the two reports.

In the case of the API report, the pre-announcement VPIN on announcement days maintains, on average, a value very similar to the value that prevails on non-announcement days (at about 16 percent), while, following the announcement, this value increases to about 19 percent and

[13]This range is similar to the grid of bucket sizes used in the big-data study by Wu et al. (2013).

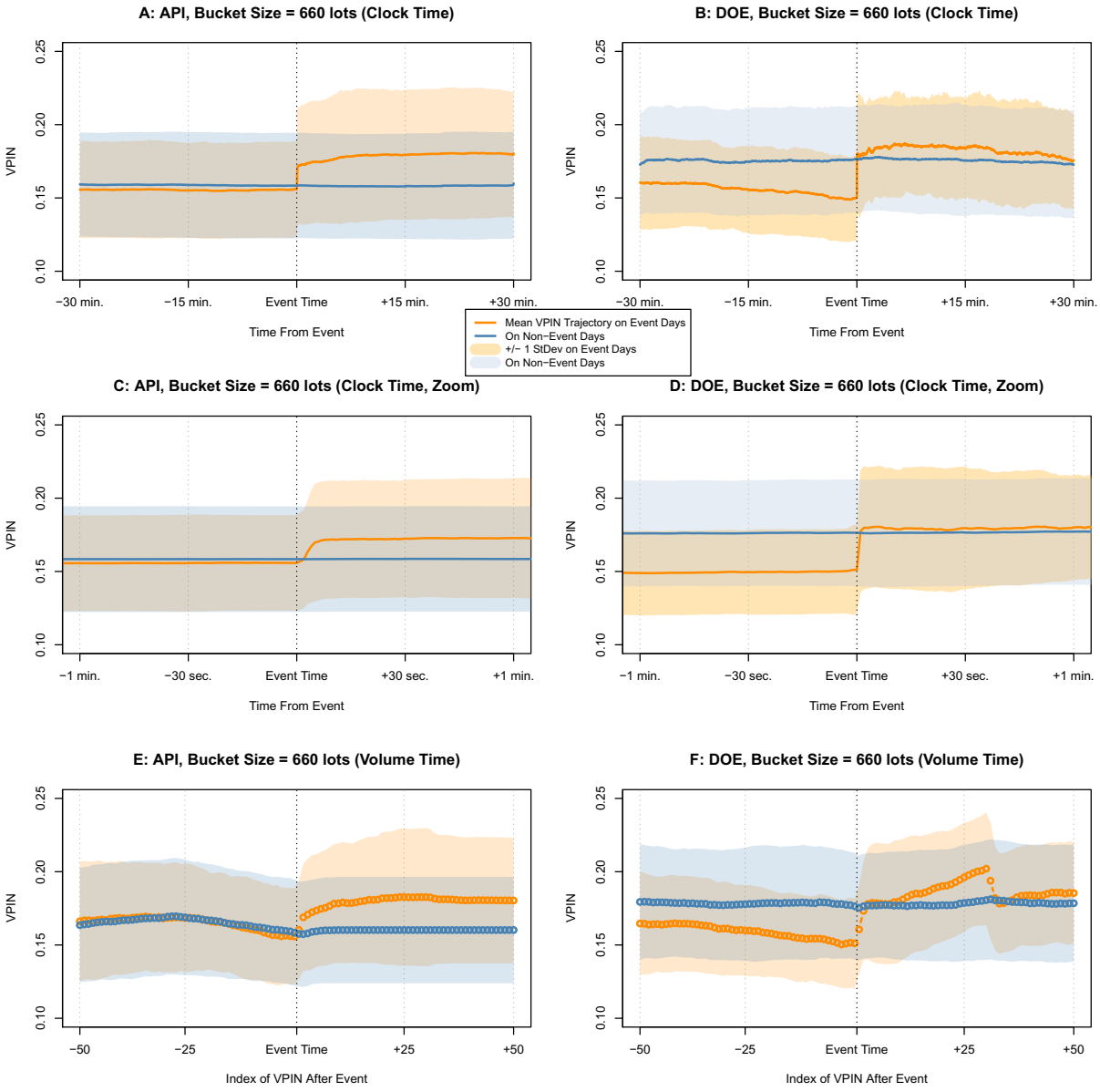


Figure 2.7: Average VPIN time series trajectories (solid lines) for the bucket size category B_1 and for both types of announcements (API - left panels, DOE - right panels). The shaded bands represent point intervals of +/- 1 standard deviation of the VPIN trajectories superimposed on the mean VPIN trajectories. Orange elements show announcement days trajectories, while blue elements represent non-announcement days during the corresponding time-of-day periods. Panels A - D show the same elements expressed in clock-time. The time window includes 30 minutes before and after each announcement and a zoomed-in version including one minute before and after each announcement. Panels E - F provide an alternative perspective by displaying the results represented in volume-time, including the 50 VPIN values observed immediately before and the 50 VPIN values observed immediately after each announcement.

remains at this level until the end of the period. Interestingly, in the case of the DOE report, the last pre-announcement VPIN tends to be, on average, 3 percentage points smaller when compared to the non-announcement days informed trading activity. An approximate 1 percentage point drop in the VPIN value occurring during the last 15 minutes before the DOE report release is also notable. Indeed, during the entire 30 minute pre-announcement time window considered, we observe systematically smaller (with a slight decreasing trend) average levels of informed trading probability, implying a greater presence of uninformed (liquidity) traders during these time periods. We reiterate that one possible reason for such behaviour by liquidity traders may be their desire to reduce their crude oil price exposure ahead of the expected announcement event by hedging, reducing, or closing their open positions in WTI futures contracts. Following the DOE announcement, an approximately 4 percentage point surge in VPIN tends to occur, reaching the average value of 18 percent roughly 10 minutes after the announcement.

Additional insights appear in the zoomed-in versions of these charts. In particular, while the most substantial increase in VPIN values occurs for the API report, on average, within the first 10 seconds after the announcement, the VPIN trajectory for the DOE report surges within the very first second after the DOE report figures are published, suggesting a considerable presence of algorithmic trading reacting to newly published information with very low latency. We note, however, that these differences are partly caused by the slower updating of the VPIN metric due to the relatively lower average volume traded during API announcements.

In order to account for clock-time and volume-time interpretation discrepancies, Panels E and F offer a volume-time interpretation, as a unit movement along the horizontal axis represents a constant amount of volume traded (one volume bucket). Considering these panels, we do observe different increase trajectories between the two reports. For the API report, the average VPIN increases by about 1 percentage point between the first and the second post-announcement volume bucket and then increases gradually by about 2 additional percentage points over the course of the subsequent 25 volume buckets, while, for the DOE report, the most dramatic increase in VPIN of about 3 percentage points occurs during the first two post-announcement volume buckets. Subsequently, the average VPIN trajectory continues to increase, as the lower (pre-announcement) volume imbalance data leave the 30 volume buckets moving average window used for the VPIN calculation. This increase continues precisely until the 30 post-announcement volume buckets become filled. Afterwards, the post-announcement average VPIN tends to stabilize at a level that is approximately 4 percentage points higher than the last pre-announcement VPIN. We conjecture that, in the case of the DOE report, the strongest reaction of informed traders occurs virtually immediately after the announcement (corresponding to the two volume buckets, or about 1,300 lots of traded volume) and is followed by a lower (although still higher than the pre-announcement level) informed order flow intensity.

We emphasize that the above descriptions refer to average VPIN trajectories. The actual

realized trajectories exhibit much noisier behaviour, as may be observed from the shaded +/- 1 standard deviation bands in Figure 2.7. Moreover, considering the almost flat (or decreasing) VPIN profiles during the pre-announcement time periods, no systematic presence of increased informed trading activity (potentially associated with insider trading) is observed. We also note that no clear relationship between the VPIN trajectory and the particular announcement surprise value is discernable from this graphical representation. We thus now proceed with further quantitative investigation of the informed trading activity dynamics around inventory announcements. With that purpose and on the basis of the above observations, we next test the three hypotheses formulated in section 2.4.

2.5.3 Hypotheses Testing

Hypothesis 1

On API (DOE) announcement days, the probability of informed trading as measured by VPIN before an API (DOE) announcement does not significantly differ from the value after the API (DOE) announcement.

We reiterate that, for a given parameter m (determining the number of volume buckets used to calculate the quantities in the table) and a given inventory report category, we formulate two testable versions of the null hypothesis in 2.8 and 2.9 as:

$$H_0^{(1a)} : E\left[VPIN_w^{m,Before} - VPIN_w^{m,After}\right] = 0$$

$$H_0^{(1b)} : \text{The distribution of } VPIN_w^{m,Before} - VPIN_w^{m,After} \text{ is symmetric around zero.}$$

We apply these tests for both types of announcements and for the following values of m: 1, 2, 3, 5, 10. Table 2.2 shows the results for the smaller bucket size category B_1.

On the basis of the p-values of the two statistical tests used in this procedure and reported in Table 2.2, we conclude that, on announcement days, the probability of informed trading as measured by VPIN before the API (DOE) announcement differs from the value observed after the API (DOE) announcement in a statistically significant manner. For both reports, the average VPIN observed shortly after the announcement is significantly greater than the one observed shortly prior to the announcement. This difference, in the case of the bucket size B_1 and for a very short window around the announcement, is approximately equal to 0.5 percentage point for the API report and 1 percentage point for the DOE report. Considering larger values of m (corresponding to a less immediate post-announcement changes), the VPIN increase equals approximately 2 and 2.5 percentage points for the API and DOE reports, respectively. Relative to the average values observed before the announcement, these increases correspond to an approximate 15 percent increases in informed trading probability for both reports.

Table 2.2: Hypothesis 1 results for the smaller bucket size category $B_1 = 660$ lots.

The six rows for each bucket size - announcement type pair show the average values of $VPIN_w^{m,Before}, VPIN_w^{m,After}$, their difference, the p-values of the Student's t-test (null hypothesis: $H_0^{(1a)} : E\left[VPIN_w^{m,Before} - VPIN_w^{m,After}\right] = 0$) and the paired Wilcoxon signed rank test (null hypothesis $H_0^{(1b)}$: the distribution of $VPIN_w^{m,Before} - VPIN_w^{m,After}$ is symmetric around zero) and the number of observations available for testing in each setting. Parameter m determines the number of volume buckets (therefore, the number of VPIN values) used to calculate the quantities in the table. For greater values of m, the quantities calculated are based on greater traded volume. The size of one volume bucket equals 660 lots.

No. of volume buckets m:			1	2	3	5	10
Bucket size B_1	API	Mean VPIN Before	0.156	0.156	0.156	0.156	0.163
		Mean VPIN After	0.160	0.165	0.167	0.169	0.183
		Mean VPIN Change	0.005	0.009	0.011	0.013	0.020
		t-test pval	0.000	0.000	0.000	0.000	0.000
		Wilcoxon pval	0.000	0.000	0.000	0.000	0.000
		No. of obs	155	155	155	155	114
	DOE	Mean VPIN Before	0.151	0.151	0.151	0.151	0.152
		Mean VPIN After	0.161	0.167	0.171	0.174	0.176
		Mean VPIN Change	0.009	0.016	0.019	0.023	0.024
		t-test pval	0.000	0.000	0.000	0.000	0.000
		Wilcoxon pval	0.000	0.000	0.000	0.000	0.000
		No. of obs	156	156	156	156	156

Overall, we **reject** Hypothesis 1 and conclude that oil inventory publications are associated with significant and rapid increases in the probability of informed trading in the WTI market. As is apparent from Table 2.2, this decision is robust with respect to the report type and to the number of volume buckets m. Moreover, as the results for the larger bucket size category B_3 (reported in Appendix 2.A) show, this decision is robust with respect to the bucket size category as well.

We note that this result is intuitively expected and confirms our observation that the US crude oil inventory level is perceived as a very important fundamental indicator in the oil market. The public release of new information causes an immediate increase in informed trading activity and is timely captured by the VPIN metric.

Hypothesis 2

a) On API announcement days, the magnitude of the VPIN change around an API announcement does not depend on the magnitude of the API surprise.

b) On DOE announcement days, the magnitude of the VPIN change around a DOE announcement does not depend on the magnitude of the DOE surprise and does not depend on the magnitude of the corresponding previous API surprise.

Table 2.3 displays the estimation results for the set of linear models defined by 2.10 and 2.11:

$$
\begin{aligned}
Y_w^{(m,API)} &= \alpha^{(m,API)} + \beta^{(m,API)} SURN_w^{API} + \gamma^{(m,API)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,API)} \\
Y_w^{(m,DOE)} &= \alpha^{(m,DOE)} + \beta^{(m,DOE)} SURN_w^{DOE} + \gamma^{(m,DOE)} \left(SURN_w^{DOE}\right)^2 \\
&\quad + \delta^{(m,DOE)} SURN_w^{API} + \rho^{(m,DOE)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,DOE)},
\end{aligned}
$$

where $Y_w^{(m,\cdot)} = VPIN_w^{m,After} - VPIN_w^{m,Before}$, as defined in section 2.4. We assess Hypothesis 2 by testing the joint significance of the non-intercept coefficients in these regressions. This procedure is equivalent to the null hypotheses defined in 2.12 and 2.13:

$$
\begin{aligned}
H_0^{(2,API)} &: \beta^{(m,API)} = \gamma^{(m,API)} = 0 \\
H_0^{(2,DOE)} &: \beta^{(m,DOE)} = \gamma^{(m,DOE)} = \delta^{(m,DOE)} = \rho^{(m,DOE)} = 0
\end{aligned}
$$

The results, reported for the smaller bucket size category B_1, strongly suggest that the magnitude of the change of informed trading intensity around both announcements is significantly related to the magnitude of the surprise associated with these announcements. The test statistics of the respective joint F-tests are significant at least at the 5% level in 8 out of 10 regressions. The γ coefficients, associated with the quadratic terms, are positive and highly statistically significant for 8 out of 10 regressions, as well. The positive signs of all the significant γ coefficients imply (again in line with our intuition) that, the greater the magnitude of the surprise (in either a positive or negative direction), the greater the increase in informed trading activity will be. These results are mostly consistent for both report categories, although in the DOE case, we note 2 regressions (associated with smaller values of m) that fail to reject $H_0^{(2,DOE)}$.

Considering the DOE results, once the DOE surprise variables are controlled for, we observe mixed results for the significance of the coefficients pertaining to the preceding API surprise (δ and ρ coefficients). While the δ coefficients (pertaining to the linear term) are mostly insignificant, the quadratic terms ρ are significant at the 5% level for one regression. We assert that these results suggest an interesting presence of a mixture effect, under which the true "surprise" of the DOE announcement depends not only on the preceding API value (the surprise of which itself depends on the DOE forecast), but also on the DOE forecast itself.

Our estimated regression equations explain up to 8 percent of the total variation for the DOE report and up to 18 percent for the API report. Finally, we note that the results are reasonably robust with respect to the range of values m considered and with respect to the two bucket sizes, B_1 and B_3 (the results for the latter appear in Appendix 2.A).

Overall, these results lead us to **reject** Hypothesis 2 for both API and DOE reports. The more unexpected (surprising) the published inventory value is, the greater increase of the order flow toxicity as measured by VPIN is observed. If we assume that the greater surprise magnitude

Table 2.3: Hypothesis 2 results for the smaller bucket size category $B_1 = 660$ lots.

The summary estimates of the regressions $Y_w^{(m,API)} = \alpha^{(m,API)} + \beta^{(m,API)} SURN_w^{API} + \gamma^{(m,API)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,API)}$ (for the API report) and $Y_w^{(m,DOE)} = \alpha^{(m,DOE)} + \beta^{(m,DOE)} SURN_w^{DOE} + \gamma^{(m,DOE)} \left(SURN_w^{DOE}\right)^2 + \delta^{(m,DOE)} SURN_w^{API} + \rho^{(m,DOE)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,DOE)}$ (for the DOE report) are displayed. The standard errors are estimated using the Newey-West estimator. The regression results that are statistically significant at the 1%, 5% and 10% levels are marked with (***), (**) and (*) signs. Parameter m determines the number of volume buckets (therefore, the number of VPIN values) used to calculate the quantities in the table. For greater values of m, the quantities calculated are based on greater traded volume. The size of one volume bucket equals 660 lots.

No. of volume buckets m:			1	2	3	5	10
Bucket size B_1	API	α	0.003***	0.005***	0.006***	0.007***	0.012***
		β	0.010	-0.092	-0.145	-0.191	-0.193
		γ	16.861**	41.120***	52.961***	66.932***	70.837***
		No. of obs	155	155	155	155	114
		Adj. R squared	0.031	0.142	0.168	0.185	0.121
		F-test statistic	3.452**	13.746***	16.545***	18.443***	8.745***
	DOE	α	0.010***	0.016***	0.019***	0.022***	0.023***
		β	-0.108	-0.056	-0.140	-0.241	-0.416*
		γ	-2.838	11.105	23.173**	40.601**	62.139**
		δ	0.025	-0.073	-0.166	-0.266*	-0.453**
		ρ	-4.282	-7.737	-12.749*	-24.328**	-29.894*
		No. of obs	156	156	156	156	156
		Adj. R squared	-0.012	-0.004	0.030	0.063	0.078
		F-test statistic	0.536	0.856	3.214**	3.626***	4.258***

indeed induces informed market participants to trade on the new information in a more intensive manner, we assert that the Hypothesis 2 results render the VPIN metric useful from the perspective of assessing not only directional changes, but also overall changes in the magnitude of the informed order flow.

Hypothesis 3

On API (DOE) announcement days, the change in the value of VPIN shortly before an API (DOE) announcement is not associated with the magnitude of the corresponding API (DOE) surprise.

We reiterate that the testable versions of the null hypotheses correspond to the following regression models and the following null hypotheses 2.15 and 2.16:

$$
\begin{aligned}
&|SURN_w^{API}| = \alpha^{(m,API)} + \beta^{(m,API)} X_w^{(m,API)} + \varepsilon_w^{(m,API)} \\
&|SURN_w^{DOE}| = \alpha^{(m,DOE)} + \beta^{(m,DOE)} X_w^{(m,DOE)} + \varepsilon_w^{(m,DOE)} \\
&H_0^{(3,API)} : \beta^{(m,API)} = 0 \\
&H_0^{(3,DOE)} : \beta^{(m,DOE)} = 0,
\end{aligned}
$$

where $X_w^{(m,\cdot)} = VPIN_w^{(Last,\cdot)} - AVGVPIN_w^{(m,\cdot)}$ is again defined in section 2.4. The regression results are reported in Table 2.4. We are thus interested in the significance of the β coefficient estimates, and after examination of Table 2.4, we state that the results do not provide strong evidence against these null hypotheses. Out of the 10 linear regressions considered for the smaller bucket size category B_1, only one regression reports statistically significant β coefficient at the 5% level of significance. This instance pertains to the DOE announcement and the VPIN moving average length parameter $m = 5$. According to the estimated β coefficient value of 0.107, a one percentage point increase in the difference between the last pre-announcement VPIN value and the moving average of the preceding 5 VPIN values is associated with an increase of the absolute value of the normalized DOE announcement surprise by 0.00107, corresponding to a surprise of 0.107% of the prevailing total level of crude oil inventories. However, this relationship, although intriguing, is not confirmed when a larger bucket size category (results reported in Appendix 2.A) is considered. Moreover, the rather small values of the adjusted R squared coefficients provide additional evidence in favor of the null hypotheses investigated.

Overall, we do not observe a stable, statistically significant relationship between changes in informed trading activity shortly before an announcement, and the magnitude of the announcement surprise. Therefore, for both types of reports, we **do not reject** Hypothesis 3. These results suggest that "last-minute" pre-announcement changes in informed trading intensity are not predictive of how unexpected (surprising) the inventories value to be released will be. As a

Table 2.4: Hypothesis 3 results for the smaller bucket size category $B_1 = 660$ lots.

The summary estimates of the regressions $|SURN_w^{API}| = \alpha^{(m,API)} + \beta^{(m,API)} X_w^{(m,API)} + \varepsilon_w^{(m,API)}$ (for the API report) and $|SURN_w^{DOE}| = \alpha^{(m,DOE)} + \beta^{(m,DOE)} X_w^{(m,DOE)} + \varepsilon_w^{(m,DOE)}$ (for the DOE report). See the main body of the text for the definitions of the regression variables. The standard errors are estimated using the Newey-West estimator. The regression results that are statistically significant at the 1%, 5% and 10% levels are marked with (***), (**) and (*) signs. Parameter m determines the number of volume buckets (therefore, the number of VPIN values) used to calculate the quantities in the table. For greater values of m, the quantities calculated are based on greater traded volume. The size of one volume bucket equals 660 lots.

No. of volume buckets m:			1	2	3	4	5
Bucket size B_1	API	α	0.008***	0.008***	0.008***	0.008***	0.008***
		β	0.019	-0.029	-0.015	0.002	0.015
		No. of obs	156	156	156	156	156
		Adj. R sq.	-0.006	-0.005	-0.006	-0.006	-0.006
		F-test statistic	0.068	0.223	0.079	0.002	0.109
	DOE	α	0.007***	0.007***	0.007***	0.007***	0.007***
		β	0.095	0.111	0.109*	0.110*	0.107**
		No. of obs	156	156	156	156	156
		Adj. R sq.	0.002	0.010	0.017	0.024	0.029
		F-test statistic	1.358	2.625	3.611*	4.861*	5.655**

potential reason for inexplicable ex-ante and ex-post rapid changes in informed trading activity could be the presence of inside information (especially last-minute leakages), the fact that we do not observe such dynamics prior to the oil inventory announcements signals no evidence of insider trading.

2.6 Conclusion

The properties of measures that aim to quantify order flow toxicity prevailing in modern electronic limit order book financial markets are of substantial interest to all market participants (especially market makers) and regulatory authorities. The aim of this study is to contribute to the existing literature by examining the VPIN metric formulated by Easley, de Prado, and O'Hara (2010) in a high frequency environment around periodically repeating macroeconomic announcements of major importance.

We chose the weekly publications of the level of US crude oil inventories of two separate agencies, the American Petroleum Institute (API) and the US Department of Energy (DOE). The weekly periodic and sequential publishing schedule of these two reports, combined with the high frequency trade-level data on the WTI futures contract at our disposal provide an interesting setting in which changes in order flow toxicity occurring at potentially very high frequencies can be investigated.

Our estimated VPIN metric dynamics around these inventory announcements suggest that the probability of informed trading as measured by VPIN increases, on average, by approximately 20% after report publications. This increase can be observed for both the API and the DOE report. Although in both cases occurring virtually immediately, the most pronounced VPIN adjustment after the DOE report publication takes place, on average, within the first two seconds, about three times faster than after the API report.

In the case of the DOE report, we additionally observe systematically smaller VPIN values prevailing during the 30 minutes leading up to the report publication time, as opposed to the same time of day on non-announcement days. We conjecture that this interesting observation may be associated with relatively greater activity of liquidity-demanding traders aiming to reduce their oil price exposure due to the upcoming announcement, as well as with relatively lower activity on the part of informed and speculative traders, willing to engage in their usual trading activity. A deeper investigation of the association of various order flow toxicity and liquidity measures (such as the potential widening of the bid-ask spreads caused by risk aversion of market makers) with this phenomenon may represent one of the possible directions of future research.

Moreover, having defined the market surprise variables based on the differences of the actual published and expected changes in the oil inventories, we document a significant relationship between the magnitude of these surprises and associated changes in order toxicity. In particular, more surprising oil inventory changes are associated with greater post-announcement percentage increases in informed trading probability, as captured by the VPIN metric.

Finally, we investigate a possible presence of last-minute insider trading activity by relating pre-announcement VPIN changes with the corresponding announcement surprise magnitudes. We do not find statistically significant evidence of the presence of such relationship.

All of our documented results appear to be considerably robust to the choices of various estimation-related parameters. We conclude that VPIN represents a potentially useful metric of adverse selection measurement in high-frequency limit order book markets.

2.A Appendix

In this section, we provide the results for another bucket size category, B_3, corresponding to 1981 WTI futures contracts. Figure 2.8 displays the average VPIN trajectories around the two announcement reports, together with the corresponding time-of-day trajectories on non-announcement days. Table 2.5, Table 2.6 and Table 2.7 show the respective results for Hypotheses 1, 2, and 3 when the VPIN values based on this larger bucket size category are considered. Overall, compared with the results in the main body of the text, we note that the results are considerably robust to the bucket size category choice.

Table 2.5: Hypothesis 1 results for the larger bucket size category $B_3 = 1981$ lots.

The six rows for each bucket size - announcement type pair show the average values of $VPIN_w^{m,Before}$, $VPIN_w^{m,After}$, their difference, the p-values of the Student's t-test (null hypothesis: $H_0^{(1a)} : E\left[VPIN_w^{m,Before} - VPIN_w^{m,After}\right] = 0$) and the paired Wilcoxon signed rank test (null hypothesis $H_0^{(1b)}$: the distribution of $VPIN_w^{m,Before} - VPIN_w^{m,After}$ is symmetric around zero) and the number of observations available for testing in each setting. Parameter m determines the number of volume buckets (therefore, the number of VPIN values) used to calculate the quantities displayed in the table. For greater values of m, the quantities calculated are based on greater traded volume. The size of one volume bucket equals 1,981 lots.

No. of volume buckets m:			1	2	3	5	10
Bucket size B_3	API	Mean VPIN Before	0.104	0.104	0.106	0.113	0.136
		Mean VPIN After	0.108	0.109	0.111	0.122	0.150
		Mean VPIN Change	0.003	0.005	0.005	0.009	0.014
		t-test pval	0.000	0.000	0.000	0.000	0.003
		Wilcoxon pval	0.000	0.000	0.000	0.000	0.005
		No. of obs	155	155	150	79	14
	DOE	Mean VPIN Before	0.105	0.105	0.105	0.106	0.107
		Mean VPIN After	0.109	0.111	0.111	0.112	0.113
		Mean VPIN Change	0.004	0.006	0.006	0.006	0.007
		t-test pval	0.000	0.000	0.000	0.000	0.000
		Wilcoxon pval	0.000	0.000	0.000	0.000	0.000
		No. of obs	156	156	156	156	156

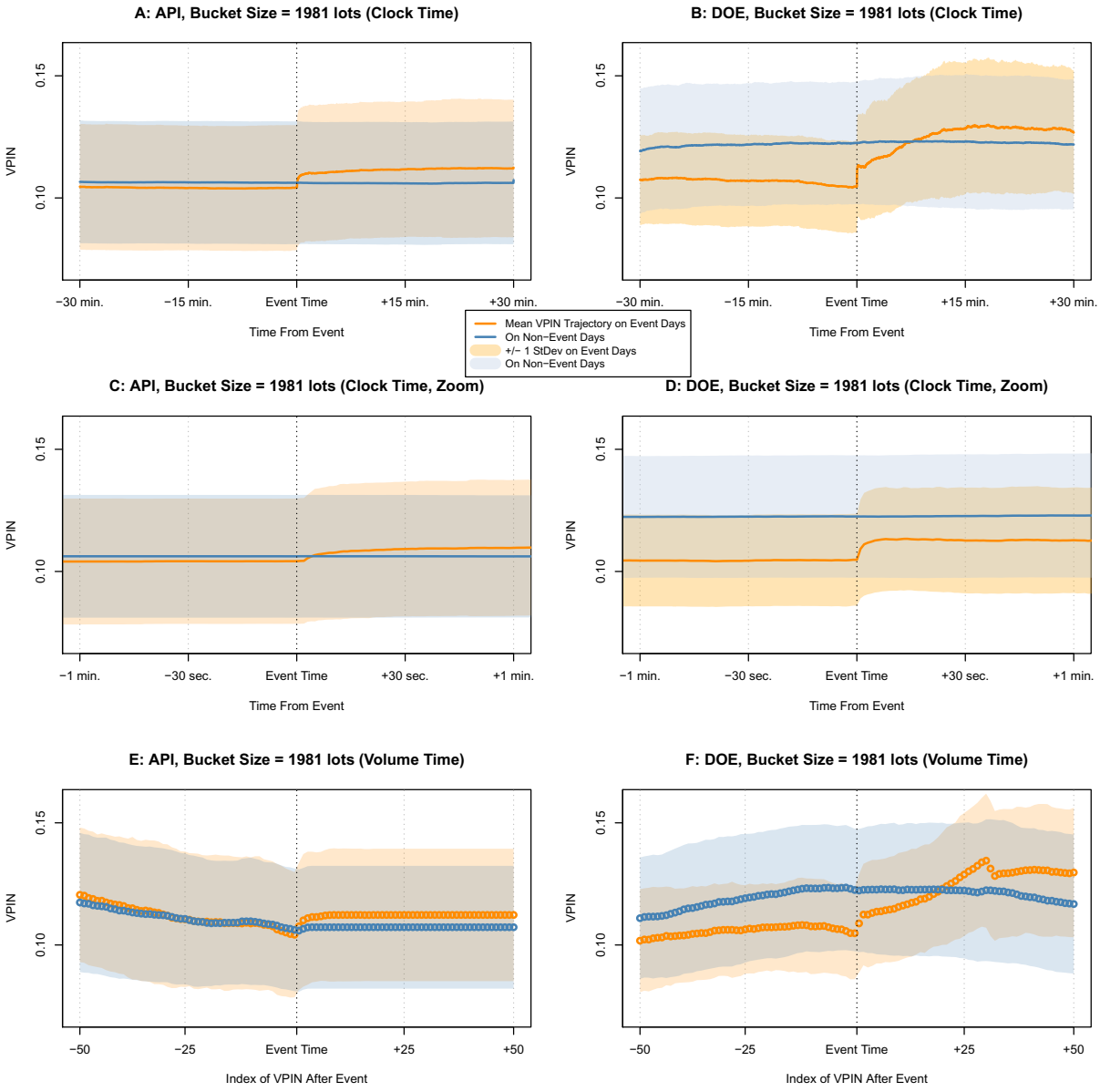


Figure 2.8: Average VPIN time series trajectories (solid lines) for the bucket size category B_3 and for both types of announcements (API - left panels, DOE - right panels). The shaded bands represent point intervals of +/- 1 standard deviation of the VPIN trajectories superimposed on the mean VPIN trajectories. Orange elements show announcement days trajectories, while blue elements represent non-announcement days during the corresponding time-of-day periods. Panels A - D show the same elements expressed in clock-time. The time window includes 30 minutes before and after each announcement and a zoomed-in version including one minute before and after each announcement. Panels E - F provide an alternative perspective by displaying the results represented in volume-time, including the 50 VPIN values observed immediately before and the 50 VPIN values observed immediately after each announcement.

Table 2.6: Hypothesis 2 results for the larger bucket size category $B_3 = 1981$ lots.

The summary estimates of the regressions $Y_w^{(m,API)} = \alpha^{(m,API)} + \beta^{(m,API)} SURN_w^{API} + \gamma^{(m,API)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,API)}$ (for the API report) and $Y_w^{(m,DOE)} = \alpha^{(m,DOE)} + \beta^{(m,DOE)} SURN_w^{DOE} + \gamma^{(m,DOE)} \left(SURN_w^{DOE}\right)^2 + \delta^{(m,DOE)} SURN_w^{API} + \rho^{(m,DOE)} \left(SURN_w^{API}\right)^2 + \varepsilon_w^{(m,DOE)}$ (for the DOE report) are displayed. The standard errors are estimated using the Newey-West estimator. The regression results that are statistically significant at the 1%, 5% and 10% levels are marked with (***), (**) and (*) signs. Parameter m determines the number of volume buckets (therefore, the number of VPIN values) used to calculate the quantities in the table. For greater values of m, the quantities calculated are based on greater traded volume. The size of one volume bucket equals 1,981 lots.

No. of volume buckets m:			1	2	3	5	10
Bucket size B_3	API	α	0.002***	0.002***	0.003***	0.006***	0.009
		β	-0.113**	-0.131**	-0.117*	-0.112	-0.351
		γ	16.217***	23.419***	23.362***	20.270***	41.366
		No. of obs	155	155	150	79	14
		Adj. R squared	0.084	0.121	0.086	0.034	-0.014
		F-test statistic	8.061***	11.619***	7.966***	2.371*	0.908
	DOE	α	0.004***	0.006***	0.006***	0.007***	0.008***
		β	0.044	-0.066	-0.096	-0.082	-0.105
		γ	17.703***	24.427***	25.806**	25.013**	17.937
		δ	0.021	-0.055	-0.084	-0.082	-0.068
		ρ	-12.234**	-20.535***	-20.858***	-20.097***	-21.891**
		No. of obs	156	156	156	156	156
		Adj. R squared	0.064	0.121	0.103	0.067	0.031
		F-test statistic	3.630***	6.327***	5.429***	3.770***	2.222*

Table 2.7: Hypothesis 3 results for the larger bucket size category $B_3 = 1981$ lots.

The summary estimates of the regressions $|SURN_w^{API}| = \alpha^{(m,API)} + \beta^{(m,API)} X_w^{(m,API)} + \varepsilon_w^{(m,API)}$ (for the API report) and $|SURN_w^{DOE}| = \alpha^{(m,DOE)} + \beta^{(m,DOE)} X_w^{(m,DOE)} + \varepsilon_w^{(m,DOE)}$ (for the DOE report). See the main body of the text for the definitions of the regression variables. The standard errors are estimated using the Newey-West estimator. The regression results that are statistically significant at the 1%, 5% and 10% levels are marked with (***), (**) and (*) signs. Parameter m determines the number of volume buckets (therefore, the number of VPIN values) used to calculate the quantities in the table. For greater values of m, the quantities calculated are based on greater traded volume. The size of one volume bucket equals 1,981 lots.

No. of volume buckets m:			1	2	3	4	5
Bucket size B_3	API	α	0.008***	0.008***	0.008***	0.008***	0.008***
		β	0.166	0.167	0.140*	0.093	0.069
		No. of obs	156	156	156	156	156
		Adj. R sq.	0.006	0.009	0.007	0.001	-0.002
		F-test statistic	1.940	2.396	2.168	1.171	0.753
	DOE	α	0.007***	0.007***	0.007***	0.007***	0.007***
		β	0.096	0.062	0.050	0.037	0.027
		No. of obs	156	156	156	156	156
		Adj. R sq.	0.001	-0.003	-0.004	-0.005	-0.006
		F-test statistic	1.083	0.521	0.406	0.251	0.147

Chapter 3

Spillovers in the Eurozone Sovereign Bond Market: Causes and Dynamics.

3.1 Introduction

One of the key properties of well-functioning and effective risk management practices in modern globalized and highly interconnected financial markets is the ability to quantify the potential for a price shock transmission among individual assets within a portfolio. Unsurprisingly, due to its growing importance from both the risk management and regulatory points of view, studies of the interconnectedness of financial markets have gained considerable attention in academic literature, from the perspective of introducing innovative concepts and of devising new metrics aiming to quantify the degree of interconnectedness (Adrian and Brunnermeier (2011), Engle and Kelly (2012), Acharya, Engle, and Richardson (2012), Diebold and Yılmaz (2014), Han et al. (2016), and Baruník and Kley (2019)) and in terms of empirically applying these concepts to a wide variety of scenarios, settings, and asset classes (Christiansen (2007), Costantini, Fragetta, and Melina (2014), Baruník, Kočenda, and Vácha (2016), Fernández-Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2016), and Tan et al. (2020)).

The aim of this study is to contribute to the empirical stream of literature on financial contagion by focusing on the European sovereign bond market. We utilize daily data on the four most liquid Eurozone sovereign bond futures markets during the time period spanning October 2015 to July 2020. The daily closing prices of futures contracts on long-term government bonds of France, Germany, Italy, and Spain are used in our analysis. Our chosen setting offers an investigation of potentially interesting spillover relationships within the European sovereign bond market, as the time period under consideration contains several geopolitical, macroeconomic, and social events of substantial regional and global significance, including the adoption of economic

policies by the Trump administration, the Italian political crisis, and the Covid-19 pandemic.

We utilize the connectedness framework proposed in the seminal study by Diebold and Yilmaz (2012) and enhanced later by Diebold and Yılmaz (2014). The authors propose several measures based on generalized forecast error variance decomposition (GFEVD) matrices of various vector autoregressive (VAR) processes. Depending on the particular type of the variables entering the underlying VAR processes used for estimation, overall magnitude and directional relationships of spillover dynamics in either log-returns or volatilities within the system can be quantified by applying various transformations to the estimated GFEVD matrices. Moreover, we provide the time development of these spillover characteristics by performing a rolling subsample analysis. Finally, we investigate potential associations of such spillover dynamics with underlying macroeconomic and geopolitical developments using panel data regression analysis.

We document the following findings. On the basis of the entire available sample, the main indicator of the overall system's connectedness (the so-called Total Spillover Index) ranges between 53 and 60 percent, based on whether log-returns or volatilities are considered the VAR process input variables. This suggests that the Eurozone bond futures market as represented by our sample constitutes a system that entails a considerable degree of risk contagion. However, when contrasted to major studies on other asset classes (such as the extreme example of 97% volatility connectedness of international financial markets documented by Diebold and Yılmaz (2014)), we state that the degree of financial contagion in the European sovereign bond market is moderate. The greatest directional contagion effect (the so-called Net Spillover Index) is observed for volatility spillovers in the "from Italy to others" direction. We argue that the turbulent political situation in Italy that was present throughout a substantial portion of the sample was associated not only with greater volatility of Italian government bond futures, but measurably contributed to increased volatility of the bond markets of the remaining major Eurozone countries as well. We deem this finding particularly interesting, as Italian bond futures consistently represented only around 10% of the aggregated system's traded volume.

However, when utilizing the rolling subsample approach and considering the results from a more detailed perspective, we state that there are several dimensions in which the characteristics of the spillover relationships differ substantially. Firstly, the degree of total connectedness exhibits considerable time variation, ranging between slightly above 30% and 72% (recorded for volatility connectedness in July, 2019 and mid-March, 2020, respectively) and increasing rapidly in response to turbulent and unexpected macroeconomic and geopolitical shocks. There are two shocks in our sample for which we observe these dynamics in the most pronounced form: the Italian political crisis (culminating on May 29, 2018) and the coronavirus outbreak volatility surge (ignited on March 12, 2020). Moreover, as the panel data regression analysis results suggest, during these shock-related crisis periods, we find evidence of a significant association between a country's Net Spillover Index and its underlying macroeconomic fundamentals (in

particular, the changes in the debt to GDP ratio and the consumer confidence indicator).

Additionally, we observe differences with respect to the choice of the VAR estimation setting, depending on whether log-returns or volatilities are considered. These differences manifest both in aggregate connectedness measures and in measures quantifying directional relationships between the variables in the system. The results for our sample of countries offer an interesting distinction between the "central" countries of Germany and France and the "peripheral" countries of Italy and Spain. Additionally, while France is documented to be the overall principal transmitter of log-return spillovers and the main receiver of volatility spillovers, opposite properties are observed for Italy, the sample's major volatility spillovers transmitter.

Finally, we document Germany's bond market as the sample's predominant receiver of spillovers in log-returns, despite being decisively the most liquid market in our study. However, we assert that Germany may represent at various time periods as strongly a role of the major transmitter of volatility spillovers as of their receiver. Germany's volatility transmitting role is most pronounced during periods of heightened global tensions, such as late 2018 and early 2019 (the main US-China trade war dispute and the Brexit deal negotiations), and the most dramatic worldwide volatility shock in decades (the Covid-19 pandemic outbreak). These observations lead us to conclude that, while the volatility transmitting role of Italy was driven mainly by its internal turbulent political development, Germany is the sample's primary volatility transmitter of the effects of major global affairs.

Overall, we assert that the connectedness measures investigated in this study produce nuanced and elaborate insights into the risk contagion structure present in financial markets. Within the context of our analysis, we assert that the dynamics of the cross-country risk contagion present in the Eurozone sovereign bond market system deserves the attention of both policymakers and market participants, particularly during episodes of increased geopolitical or macroeconomic tensions.

The remainder of this paper is structured as follows. Section 3.2 reviews relevant risk contagion studies, particularly in the context of sovereign bond markets. In section 3.3, we introduce the connectedness measures we apply. Section 3.4 describes the European bond market data available for estimation and overviews the main macroeconomic and geopolitical developments during the period. The connectedness results are described and discussed from multiple perspectives in section 3.5. Section 3.6 concludes.

3.2 Literature Review

Due to its growing importance from investment, risk management and regulatory perspectives, the study of the interconnectedness of financial markets, especially in terms of volatility spillovers, has gained considerable attention in academic literature.

Several different research strategies aiming to quantify the concept of connectedness of financial assets have been proposed. A substantial set of studies utilize and enhance the intuitive concept of correlation. Engle and Kelly (2012), for instance, introduce the equi-correlation methodology, focusing on average pairwise correlations and providing a relatively simple estimation strategy of arbitrarily large covariance matrices. Adrian and Brunnermeier (2011) define the CoVaR (conditional Value at Risk) metric, while Acharya, Engle, and Richardson (2012) propose the MES (marginal expected shortfall) metric, both quantifying the relationship between the individual riskiness of a financial asset and variations in global market fundamentals. Additional focus has been dedicated to analyzing non-average quantile-based measures, such as the cross-quantilogram approach introduced by Han et al. (2016) or the quantile coherence measure proposed by Baruník and Kley (2019).

Another way of quantifying cross-asset spillovers has been introduced and defined in the seminal study by Diebold and Yilmaz (2009) in which the metrics of volatility spillovers are based on the connectedness framework. The authors utilize the concept of forecast error variance decomposition (FEVD) of a VAR process. Since an (i, j) element of the FEVD matrix represents the fraction of the variable i's forecast error variance that may be attributed to shocks in another variable j of that same system, spillover measures based on FEVD matrices represent an intuitive way to quantify the degree and the direction of financial contagion prevailing within the system. Diebold and Yilmaz (2012) and Diebold and Yılmaz (2014) further enhance connectedness metrics by adopting the concept of generalized FEVD matrices (Koop, Pesaran, and Potter (1996), Pesaran and Shin (1998)). Introducing the generalized FEVD matrices accounts for the inherent property of the original FEVD matrices not being robust to the ordering of variables in the VAR system.

Since its inception, connectedness methodology has been applied to numerous research settings, an exhaustive account of which is beyond the scope of this section. Nevertheless, we aim to provide a representative overview. Among the studies analyzing stock market performance, Diebold and Yilmaz (2015) characterise return and volatility spillovers across a universe of major European and American financial institutions, utilizing 10 years of daily data. While they document a strong unilateral direction of spillovers from America to Europe during the 2007 - 2008 financial crisis. starting in late 2008, the relationship becomes increasingly bilateral, especially during the onset of the European sovereign debt crisis. Jiang et al. (2020) investigate the volatility spillovers at different frequencies across different industry sectors in China. The authors state that the main direction of volatility spillovers in China occurs from the industrial and construction sectors towards the banking and real estate sectors.

Another important universe of assets analyzed within the connectedness framework is the forex market. Diebold and Yılmaz (2015) study nine major currency pairs, documenting a moderate post-2007 crisis increase in their overall connectedness. Other studies analyzing the

connectedness of the forex market include Antonakakis (2012), Greenwood-Nimmo, Nguyen, and Rafferty (2016), and Baruník, Kočenda, and Vácha (2017). Do, Brooks, and Treepongkaruna (2015) examine a universe of assets including the stock and forex markets. Employing high frequency data, the authors find positive spillover effects between these two markets, occuring in both directions, especially in turbulent periods. Finally, Yi, Xu, and Wang (2018) examine the popular universe of cryptocurrencies. Among the 52 cryptocurrencies they analyze, they find that Bitcoin is an important, albeit not a dominant net transmitter of connectedness.

Among studies focusing on risk contagion in energy markets, Zhang (2017) applies connectedness framework to investigate the relationships between shocks to the world's oil market and six globally important equity indices. Although the importance of oil market spillovers is documented as being overall relatively weak, the author concludes that, during the periods of extraordinary oil market volatility, risk contagion in the direction to the equity indices becomes significant. Ji, Zhang, and Geng (2018) analyze the energy market universe, including crude oil, carbon emissions, coal, clean energy, and electricity markets, and find that the oil market is dominant in terms of both return and volatility spillovers transmission. Conversely, the electricity market is shown to be the main net spillover receiver. Connectedness of the carbon emissions market to energy and financial markets is studied by Tan et al. (2020). The authors document a substantial time variation of all of the network connectedness measures, also driven by the presence of a structural break in the spillover effect of the carbon emissions market.

Several studies enhance and refine the connectedness methodology to identify more elaborate aspects of financial contagion. For example, Baruník, Kočenda, and Vácha (2016) decompose volatility into "good" and "bad" volatility, applying the concept of the realised semi-variance of Barndorff-Nielsen, Kinnebrock, and Shephard (2008) to US stock market data. Subsequently, utilising high-frequency intraday prices of the most liquid forex futures contracts, Baruník, Kočenda, and Vácha (2017) find that the volatility spillovers associated with "bad" volatility (negative shocks) dominate "good" volatility spillovers. While the positive spillovers are associated with the 2007 subprime crisis, the negative spillovers are mostly tied to the European sovereign debt crisis. Antonakakis, Gabauer, and Gupta (2019) combine the connectedness framework with the Bayesian time-varying parameter vector autoregressions and measure the spillover time dynamics of four developed stock markets without the need for rolling subsamples. While they find that the US and EU stock markets are major transmitters of monetary policy spillovers, the UK and Japanese markets are mostly their net receivers. Finta and Aboura (2020) examine higher moments spillovers among the world's most liquid stock markets, focusing not only on volatility, but also on skewness risk premium spillovers. The authors find that these two kinds of risk premia exhibit substantial bi-directional spillovers.

In line with the focus of this study, we next overview the research on financial contagion within the universe of sovereign bond markets. Earlier studies document mostly relatively strong

common determinants of either global or regional co-movements of interest rates. Ilmanen (1995) investigates the predictable component of time-varying expected returns in six long-term government bond markets. They find that a small set of global instruments forecasts up to 12 percent of bond returns' monthly variation. Clare and Lekkos (2000) examine the relationship between the US, UK and German sovereign bond markets, and find the existence of common international factors influencing all three government yield curves, especially during periods of financial distress. Laopodis (2004) provides empirical evidence of the impact of globalization on greater synchronization of interest rates and examines the implications of this phenomenon on the conduct of monetary policy. This convergence in interest rate dynamics is particularly observed for the European Union, though the special status of Germany is maintained throughout the sample, as acknowledged by the author. In terms of bond volatility spillovers, Christiansen (2007) applies a volatility spillover model based on a GARCH framework to investigate contagion channels between the aggregate US and EU bond markets and bond markets of individual European countries. Existence of strong spillover effects towards the individual markets is documented. Furthermore, while the US market is a strong volatility transmitter to the non-Eurozone European countries, its effect on the Eurozone markets appears to be substantially weaker.

Balli (2009) investigates the degree of European sovereign bond market integration and concludes that, in the Eurozone, various macroeconomic and fiscal factors cannot explain the variation in sovereign bond yields in the monetary union era, making a case for almost perfect financial integration of these markets. However, the author observes post-global shock discrepancies among Eurozone bond markets. Other studies examining possible determinants of Eurozone's sovereign bond yield spreads have been carried out by Favero and Missale (2012) and Costantini, Fragetta, and Melina (2014). In the latter study, the authors find that the different fiscal imbalances in the form of debt-to-GDP ratios explain the variation in the corresponding sovereign yield spreads to a considerable degree. Sibbertsen, Wegener, and Basse (2014) examine the persistence of French, Italian, and Spanish bond yield spreads (against German interest rates) and find the existence of structural breaks in the persistence of these spreads occurring between 2006 and 2008. They document an evolution towards stronger persistence of the yield spreads, arguably due to the Eurozone debt crisis.

The following studies apply the connectedness methodology based on FEVD matrices to bond markets and can therefore be considered the most closely related to the aim of our study. Antonakakis and Vergos (2013) investigate more than 5 years of daily data on 9 Eurozone sovereign bond yields. Spillovers originating from the destabilizing shocks in periphery countries are observed to be more prevalent than those from core countries. Moreover, stronger within-effects of spillovers are documented for the periphery countries. Claeys and Vašíček (2014) use a factor-augmented VAR framework to examine 16 European Union bond markets. Fernández-

Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2015) study eleven European bond markets, both from the core and from the periphery of the region, using almost 15 years worth of daily bond yield data covering the subprime mortgage crisis and the European sovereign bond crisis. The authors document that approximately 55% of the forecast error variance can be attributed to cross-country shocks, while Italy and Spain are the prevalent net transmitters of shocks to bond yields, and Portugal and Ireland are their primary net receivers. Moreover, the authors find that the key macroeconomic determinants of connectedness differ between core and periphery countries.

More recently, Fernández-Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2016) extend their study of net pairwise connectedness and conclude that, following the sovereign debt crisis, the possibility of a volatility contagion from periphery countries to core countries increased considerably. Ahmad, Mishra, and Daly (2018) analyze almost 20 years of weekly bond market data of BRICS countries, together with US, European and Japanese bond market indices. The authors find that the most active net transmitters of volatility are South Africa, Russia, and Brazil, which also show greater pairwise connectedness measures relative to the rest of the sample. Finally, Hamill et al. (2021) investigate the connectedness dynamics of eleven Eurozone sovereign bond markets from 2005 to 2011, focusing particularly on the potential indicative ability of connectedness with respect to the global financial crisis and the European debt crisis that followed. The authors document the tendency of the system's connectedness to decrease with the onset of these crises, with peripheral countries approaching almost complete isolation from the rest of the system.

3.3 Methodology

We consider an N-dimensional time series, observed at daily frequency and indexed by the "day" index d, $X_d = (X_{1d}, \ldots, X_{Nd})', d = 1, \ldots, D$. We assume that X_d is governed by a covariance-stationary vector autoregressive process of order p:

$$X_d = \sum_{i=1}^{p} \Phi_i X_{d-i} + \varepsilon_d, \tag{3.1}$$

with the iid disturbance vector $\varepsilon_d \sim N(0, \Sigma_\varepsilon)$ and the $(N \times N)$ AR coefficient matrices Φ_i. This VAR structure implies the existence of contenporaneous and intertemporal dependencies among the various pairs of variables in the system, incorporated in the particular elements of the coefficient matrices. We are primarily interested in quantifying the strength and directions of these dependencies, especially in the context of financial contagion emanating from innovation (shock) spillovers. Following that objective, we express the VAR process 3.1 by its corresponding

moving average (MA) representation:

$$X_d = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{d-i}, \tag{3.2}$$

where the following recursive relationships hold for the $(N \times N)$ MA coefficient matrices Ψ_i (0_N and I_N denote the $(N \times N)$ zero and identity matrix, respectively):

$$\Psi_i = \begin{cases} 0_N, & i < 0 \\ I_N, & i = 0 \\ \sum_{j=1}^{p} \Phi_j \Psi_{i-j}, & i > 0 \end{cases} \tag{3.3}$$

Diebold and Yılmaz (2014) calculate the spillover index based on the H-steps-ahead generalized forecast error variance decomposition (GFEVD) matrix introduced by Koop, Pesaran, and Potter (1996) and Pesaran and Shin (1998). Ω^H denotes the H-steps-ahead GFEVD matrix of the above process 3.2. For a given forecast horizon H, the (i,j)-th element of Ω^H is equal to:

$$\omega_{ij}^H = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} \left(e_i^T \Psi_h \Sigma_\varepsilon e_j\right)^2}{\sum_{h=0}^{H-1} \left(e_i^T \Psi_h \Sigma_\varepsilon \Psi_h^T e_i\right)}, i, j = 1, \ldots, N, \tag{3.4}$$

where $\sigma_{jj}, j = 1, \ldots, N$ are the diagonal elements of the non-orthogonalized disturbance covariance matrix Σ_ε and e_i is an N-dimensional column vector with the i-th element equal to one and zero otherwise.

Unlike the usual FEVD matrices, which assume orthogonal shocks and are dependent on the ordering of the variables entering the VAR process, the GFEVD matrices do not necessarily implicitly correspond to orthogonal shocks. Therefore, the row-wise sums of the elements of these matrices are generally not equal to one. By normalizing each element of Ω^H by its corresponding row-wise sum, we obtain the H-steps-ahead $N \times N$ connectedness matrix C^H with the (i,j)-th element equal to:

$$c_{ij}^H = \frac{\omega_{ij}^H}{\sum_{j=1}^{N} \omega_{ij}^H}. \tag{3.5}$$

The (i,j)-th element of the normalized GFEVD matrix C^H can be therefore interpreted as the fraction of the H-steps-ahead forecast error variance of the i-th variable that arises due to shocks to the j-th variable of the system. In this sense, the elements of this connectedness matrix operate as measures of the overall degree of the connectedness of variables within the system, as well as of the direction and strength of potential spillover effects among these variables. Note that, due to normalization, the row-wise sum of the elements of matrix C^H is always equal to 1 and the overall sum of all elements of matrix C^H is equal to N. Based on the connectedness

matrix C^H, the following spillover quantities are then readily defined (Diebold and Yılmaz (2014)):

- **Total Spillover Index** (ρ^H_{TOT}): The fraction of the aggregated forecast error variance (across all variables) that can be attributed to shock spillovers between variables and not to shocks to the variables themselves. This quantity is therefore equal to the normalized sum of all the off-diagonal elements of the connectedness matrix. Hence, its value is always between zero and one:
$$\rho^H_{TOT} = \frac{100 \cdot \sum_{i,j=1, i \neq j}^{N} c^H_{ij}}{N} \tag{3.6}$$

- **Directional Spillover Index "From"** ($\rho^H_{i\leftarrow\cdot}$): The fraction of the variable's i forecast error variance that can be attributed to shocks to all other variables in the system (except the variable i itself):
$$\rho^H_{i\leftarrow\cdot} = \sum_{j=1, i \neq j}^{N} c^H_{ij} \tag{3.7}$$

- **Directional Spillover Index "To"** ($\rho^H_{i\rightarrow\cdot}$): The amount determining how much shocks to variable i contribute to the forecast error variances of all the other variables in the system (except to the variable i itself):
$$\rho^H_{i\rightarrow\cdot} = \sum_{j=1, i \neq j}^{N} c^H_{ji} \tag{3.8}$$

- **Net Spillover Index** (ρ^H_i): The net amount of spillovers transmitted by variable i, defined as the difference between the variable i's "To" and "From" Directional Spillover Indices:
$$\rho^H_i = \rho^H_{i\rightarrow\cdot} - \rho^H_{i\leftarrow\cdot} \tag{3.9}$$

- **Own Spillover Index** c^H_{ii}: These are the diagonal elements of the connectedness matrix C^H. Each element c^H_{ii} represents the fraction of the forecast error variance of variable i arising due to shocks to variable i itself. The value of the Own Spillover Index can be interpreted as the variable i's degree of isolation, because the greater the Own Spillover Index, the smaller the values of the matrix C^H row-i's off-diagonal entries, which represent the cross-variable connectedness.

For a given N-dimensional VAR system estimate (based on a given sample of data) and a given forecast horizon H, we observe one $(N \times N)$ connectedness matrix C^H. This matrix identifies one Total Spillover Index ρ^H_{TOT}, quantifying the overall connectedness of the system. Additionally, there are N Directional Spillover Indices "From" ($\rho^H_{i\leftarrow\cdot}, i = 1, \ldots, N$), as well as N Directional

Spillover Indices "To" ($\rho^H_{i\rightarrow .}, i = 1, \ldots, N$), one of each for each of the variables in the system. Therefore, there also exist N Net Spillover Indices ($\rho^H_i, i = 1, \ldots, N$), which can be interpreted as the difference between the total amount of spillovers variable i contributes to the system and the total amount of spillovers this variable receives from the other variables in the system. Variables with positive (negative) values of the Net Spillover Indices act as the system's net transmitters (net receivers) of spillovers.

In practice, we estimate the above connectedness quantities on a data sample of a predetermined length. We index the day-d estimation sample of length D by time indices $d-D, d-D+1, \ldots, d-1, d$. To assess the statistical significance of the estimated connectedness measures, we apply the non-parametric bootstrap methodology as outlined in Diebold and Yılmaz (2015). We describe the procedure as follows:

1. For a given sample and a given estimation of the VAR process, we record the estimation residuals. We obtain a time series of D N-dimensional residual vectors.

2. One bootstrap repetition involves:

   (a) Drawing a random sample of size D with replacement from the original time series of residual vectors.

   (b) Utilizing this bootstrapped residual series and the coefficient estimates of the original VAR estimation in order to generate a new bootstrapped vector time series.

   (c) Performing a VAR estimation on this new bootstrapped vector time series.

   (d) Calculating and recording the bootstrapped GFEVD matrix as well as the implied bootstrapped connectedness matrix and the corresponding spillover measures.

3. Performing B bootstrap repetitions yields B connectedness matrices and B spillover measures of each kind. We use this bootstrapped sample to calculate the bootstrap standard errors for each spillover measure in isolation.

Within the scope of our literature review, we note that, with the exception of the Diebold and Yılmaz (2015) textbook, we did not find any connectedness study containing a similar discussion on the statistical significance of the obtained results.

Additionally, since we are also interested in observing the dynamics of the evolution of financial contagion within our set of assets, we include a rolling-subsample methodology in our estimations, in which, for each period, we fix the length D of the preceding data sample used for estimation. For a given day $d > D$, this yields the day-d estimated connectedness matrix as well as all the spillover measures. We then iterate the rolling-subsample dataset one period forward by dropping the data for the most distant past day and adding the data for the next day $(d+1)$. The estimation procedure is then applied to this altered subsample, yielding the updated

estimate of the connectedness matrix (and that of all the corresponding spillover measures) for day $(d+1)$.

To apply the methodology in practice, several choices regarding the nature of the data and the hyperparameter values need to be made:

- **Variables entering the system:** The original VAR system is based on an N-dimensional time series process X_d. We are interested in analyzing the log-return spillovers, both in terms of their overall level and of their volatility. Therefore, we estimate the VAR processes for the following three classes of variables (and denote these variables as $Y_{i,d}^{(\cdot)}$:

  - **Log-Returns:** Denoting the day-d closing price of asset i by $P_{i,d}$, we set:

$$Y_{i,d}^{(1)} = r_{i,d} = \log(P_{i,d}) - \log(P_{i,d-1}). \tag{3.10}$$

  - **Volatilities:** In our study, the second choice of the variable entering the VAR process corresponds to the day-d estimated value of the asset i's historical volatility of daily log-returns, calculated as:

$$Y_{i,d}^{(2)} = sd_{i,d} = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(r_{i,d-k} - \frac{1}{K}\sum_{k=1}^{K} r_{i,d-k}\right)^2}. \tag{3.11}$$

    We set the value of K equal to 30. We also experimented with different reasonable values of K and find that the results are robust with respect to the choice of this parameter.

  - **Log-Volatilities:** For robustness purposes, as our third choice of the VAR variables, we consider the logarithms of the above calculated volatilities. We made this choice to approximately accommodate the normality assumption of the GFEVD matrix estimation:

$$Y_{i,d}^{(3)} = \log(sd_{i,d}). \tag{3.12}$$

- **Order of the VAR process p:** in line with the well-established approach in existing literature,[1] we choose the order of the estimated VAR process according to information criteria. For each VAR estimation in our study, we set p equal to the arithmetic average (rounded to the nearest natural number) of the optimal lag lengths suggested by the Akaike, Schwarz and Hannah-Quinn Information criteria. Nevertheless, we also check the robustness of the rolling-window results for the Total Spillover Index with respect to the choice of the lag order. The robustness results are reported in Appendix 3.A.

[1] See, for example, Zhang (2017), among many others.

- **Length of the estimation window D:** We set $D = 252$, that is, for each day, the forecast error variance decomposition matrix is estimated on a sample covering the preceding 252 business days, corresponding to approximately one calendar year.

- **Forecast error horizon H:** We set the value of H equal to 20, that is, we are analyzing the error variance decomposition of the estimated model forecasts 20 business days (approximately one calendar month) into the future. To examine the robustness of the calculations, we also experiment with the values $H = 10, H = 30$. The robustness results are reported in Appendix 3.A.

3.4 Data & History Overview

We analyze daily closing prices of futures contracts on government bonds of four Eurozone countries: France, Germany, Italy, and Spain. More formally, we focus on futures contracts delivering at maturity notional long-term debt instruments (the practitioners' abbreviations of which are stated in parentheses) issued by the Republic of France (the OATs), the Federal Republic of Germany (the BUNDs), the Republic of Italy (the BTPs) and the Kingdom of Spain (the BONOs).[2] The futures contracts considered in this study are quoted at the Eurex electronic exchange and the data is sourced from Bloomberg. The Eurex Product IDs are FOAT, FGBL, FBTP and FBON, for the French, German, Italian, and Spanish bond futures. We refer to these contracts by the country names, and report the results in corresponding alphabetical order. Our daily data covers the period starting on October 26, 2015 (the first day of trading of the Spanish BONO futures contracts) and ending on July 31, 2020, yielding 1,212 daily observations in total.

According to the Eurex official contract specifications[3], the notional amount of one futures contract is equal to 100,000 EUR. Moreover, at any point in time, the underlying government bonds have remaining terms to maturity ranging from 8 to 11 years, and are paying coupons of 6% per year. At any point in time, there are three futures contracts available for each of the four underlying securities, maturing in the three nearest quarterly months of the March, June, September, and December cycles. We collect daily price and volume data for all maturities available on any given day. Subsequently, for each security and each day, the futures contract with the largest daily volume traded is considered in this study. We note that this selection procedure yields in all cases a "monotonic" progression of maturities, that is, for any two consecutive trading days, the corresponding futures contracts have either the same maturity, or the maturity date moves into the future for the second day's contract and afterwards, the preceding

[2]The abbreviations of these long-term sovereign bonds have the following origins: OAT - Obligation assimilable du Trésor, BUND - Bundesanleihen, BTP - Buoni del Tesoro Poliannuali, BONO - Bonos del Estado.

[3]`https://www.eurexchange.com/exchange-en/markets/int/fix/government-bonds`

maturity date is never observed again.

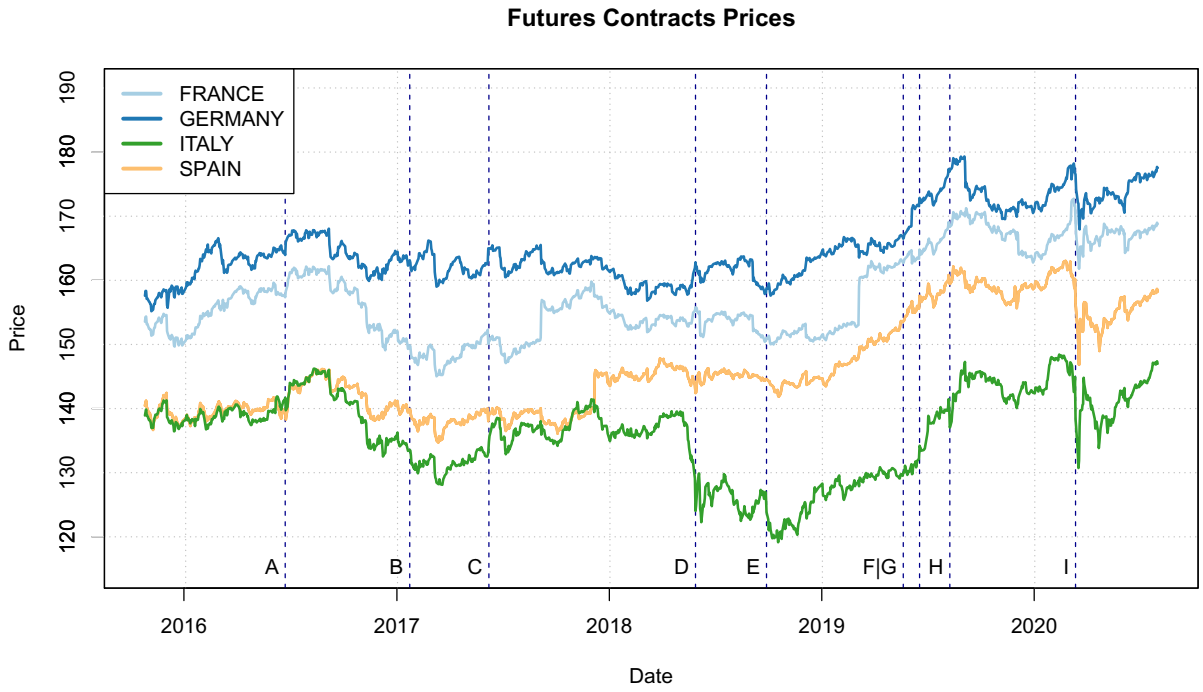


Figure 3.1: Time development of prices of the most liquid (in terms of daily volume traded) futures contracts on long-term government bonds of the four Eurozone countries considered in this study. The vertical labeled lines highlight selected political and macroeconomic events discussed in the main body of the text. Time period observed: October 26, 2015 - July 31, 2020. Source: Bloomberg.

We display the price history of the most liquid futures contract for each of the four countries in Figure 3.1. Additionally, Figure 3.2 complements the price data by displaying the time development of daily yields to maturity (YTM) for the on-the-run underlying government bonds during the same period.[4] Not surprisingly, the price and yield series move largely in unison for all four countries, reflecting the fundamental commonalities of the corresponding four economies, including their common Eurozone membership. Moreover, starting from the second half of 2016, at any day d. the prices of these government bond futures satisfy the following set of inequalities: $P_d^{GER} > P_d^{FRA} > P_d^{SPA} > P_d^{ITA}$, with the corresponding yield series satisfying $YTM_d^{GER} < YTM_d^{FRA} < YTM_d^{SPA} < YTM_d^{ITA}$. These inequalities and the dynamic nature of the corresponding yield spreads arguably primarily reflect the different prevailing levels of sovereign credit risk across the four countries. Our dataset offers an interesting opportunity to study EU fixed income market connectedness, as several stylized features, along with interesting events with considerable market impact during this period (highlighted with the "A" - "I"

[4]The yield to maturity data is sourced from Bloomberg as well.

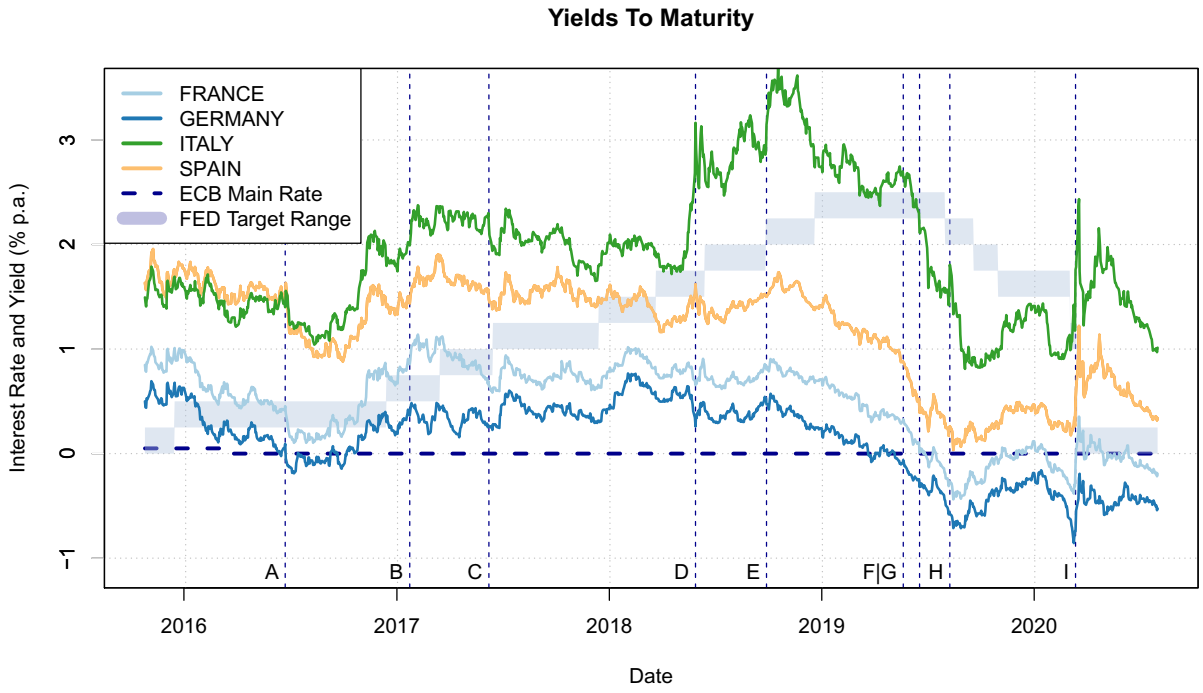


Figure 3.2: Time development of yields to maturity (expressed in percentage per annum) of the long-term government bonds of the four Eurozone countries considered in this study. The vertical labeled lines highlight selected political and macroeconomic events discussed in the main body of the text. Time period observed: October 26, 2015 - July 31, 2020. Source: Bloomberg.

letter-marked vertical lines in the Figures whenever applicable) are observable, a brief overview of which we now present.

While several global and regional events (the negative interest rate adoption by the Bank of Japan at the end of January, the surprising results of the June 23 Brexit referendum (A) and bad loans plaguing some of the largest European banks) were driving bond prices higher during the first nine months of 2016, a negative trend followed, exacerbated by the positive market perception of the election of Donald Trump, the resulting optimism regarding tax cuts (B) and the OPEC oil cuts announced in November.[5]

Perhaps the most persistent source of price volatility in the bond futures considered in this study, prevailing throughout a major segment of our sample, was the turbulent political development in Italy, the Eurozone's third largest economy (after Germany and France) and the second largest public debtor (after Greece). At the beginning of December 2016, the reforms-rejecting results of the Italian constitutional referendum were announced, triggering the resignation of the country's prime minister, Matteo Renzi. Among the first political shocks during the government led by his successor, Paolo Gentiloni, was the collapse of an electoral reform accord (originally

[5] `https://www.ft.com/content/6d24125c-c066-11e6-9bca-2b93a6856354`

agreed between the main political parties) that occurred on June 8, 2017 (C), leading multiple party leaders to call for immediate snap elections. Eventually, Gentiloni's government remained in place until the March 2018 general election. As this election resulted in a hung parliament and negotiations on forming a government lasted for several months, a substantial change in the level of both Italy's and Spain's sovereign credit risk occurred during the second quarter of 2018, reflecting the increasing market fears of the prospect of another election in Italy, the possibility of eurosceptic parties strengthening their position in the Italian government, and a vote of no confidence against the Spanish government. During this period, the Italian yields almost doubled, increasing from 1.8 % p.a. to approximately 3.4 % p.a (during a particularly volatile session on May 29 (D), the Italian BTP futures price dropped by 4.6%), while the Spanish bond yields increased by approximately one third (from 1.3 % p.a. to 1.7 % p.a.).[6] The "safe haven" French and German bonds, however, experienced a moderate decline in yields during the same period.

The Italian yields continued an upward trend in a very volatile fashion until the end of November, 2018, as the turmoil in the EU-Italy relationship regarding the country's budget deficit levels continued. Another particularly volatile period of several weeks was initiated on September 29 (E), with the publication of the news that the Italian government intended to run deficits as large as 2.4% of GDP, as opposed to the original planned 0.8%.[7] By the end of 2018, however, prevailing global fears about a potential US-China trade war dispute, and decreasing worries about Eurozone stability (with the Italian government's eventual willingness to negotiate deficit levels with Brussels) and the increasing uncertainty regarding the post-Brexit EU-UK relationship contributed to flight to safety in the form of all EU government bonds, and resulted in a gradual downward trend in the yields of all four countries, with the BTP-BUND spread narrowing by almost 50% from its October 2018 peak of 3.4 percentage points.

2019 was a year of overall monetary policy-easing rhetoric, proclaimed by both the Federal Reserve (while in 2018, the FED carried out 4 hikes of its Federal Funds target range, starting in 2019, rate hikes were put on hold and a rate cut sequence was initiated in the summer) and by the ECB (which, for example, signalled on June 18 (G) a fresh round of monetary stimulus in June[8]). In combination with the Brexit negotiation uncertainty (manifesting, for example, on May 29 (F), when the British Prime Minister Theresa May's Brexit deal proposal was rejected by members of parliament across the political spectrum[9]), and the prospect of global currency wars

[6] https://www.theguardian.com/business/live/2018/may/29/markets-eurozone-italian-spanish-political-turmoil-business-live

[7] https://www.theguardian.com/business/2018/sep/28/italy-budget-drama-all-you-need-to-know

[8] https://www.bloomberg.com/news/articles/2019-06-18/draghi-says-further-interest-rate-cuts-remain-part-of-ecb-tools-jx1j9wch

[9] https://www.theguardian.com/politics/2019/may/21/may-offers-mps-vote-on-second-referendum-in-new-brexit-deal

(with the Chinese Renminbi depreciating against the US dollar to a 10-year low), the bond yields continued the downward trend initiated in late 2018, resulting in a record $17 trillion worth of debt trading in the negative yield territory by the end of August 2019 (including the German and the French long-term sovereign debt).[10] This negative trend, however, reversed during the final months of 2019. Among the four countries in our sample, the Italian BTP bonds were the ones exhibiting the greatest volatility throughout the year, notably on August 9 (H), as the no confidence motion announced by Deputy Prime Minister Salvini against Prime Minister Conte resulted in government coallition collapse. Subsequently, government-forming negotiations between the opposition Democratic Party and the anti-establishment 5-Star Movement unfolded, and a new government (however, again under Conte's leadership) won confidence in September.[11]

More recently, during the coronavirus outbreak at the beginning of 2020, a sharp increase in the yields and a surge in the futures price volatility can be observed (I). With the gradual resolution of the uncertainty regarding both the outbreak's severity and the monetary and fiscal packages aimed at attenuating the negative shocks associated with the rapid and extensive lockdown measures, all four bond yields gradually returned to their pre-outbreak levels by the end of July, 2020. [12]

Apart from the price and yield data, the traded volume data offer interesting insights into the Eurozone sovereign fixed income futures market as well. Figure 3.3 shows the aggregate number of the most liquid futures contracts traded each day (the left panel, 5-day moving average) and the share of this aggregated volume that can be attributed to each of the four countries considered (the right panel, 15-day moving average). On average, approximately 900,000 contracts were traded each day, with pronounced quarterly peaks surrounding the futures contracts maturities in March, June, September, and December cycles. As the maturity of a futures contract approaches, investors interested in maintaining their positions need to close their positions in the maturing contracts and open the corresponding positions in the subsequent maturities, a procedure known as the "roll-over". During these periods, it was not uncommon to observe more than 1.5 million daily contracts being traded on aggregate.

Out of this aggregate daily volume, the share of volume that can be attributed to German futures is always greater than 70% . The French OATs represent approximately 15% of the aggregate daily volume (with a slight increasing trend in this proportion) and the Italian BTPs represent approximately 10%. Overall, the share of the Spanish BONO futures is always smaller than 0.5%, and is only marginally visually discernible in Figure 3.3. There are relatively stable

[10] https://www.ft.com/content/7ddaf4b6-d8e2-11e9-8f9b-77216ebe1f17

[11] https://www.reuters.com/article/us-italy-politics-zingaretti/italy-pd-party-tells-president-ready-to-form-government-with-5-star-idUSKCN1VI1SJ?il=0

[12] https://www.ecb.europa.eu/press/pr/date/2020/html/ecb.pr200318_1~3949d6f266.en.html
https://www.cnbc.com/2020/04/13/coronavirus-update-here-is-everything-the-fed-has-done-to-save-the-economy.html

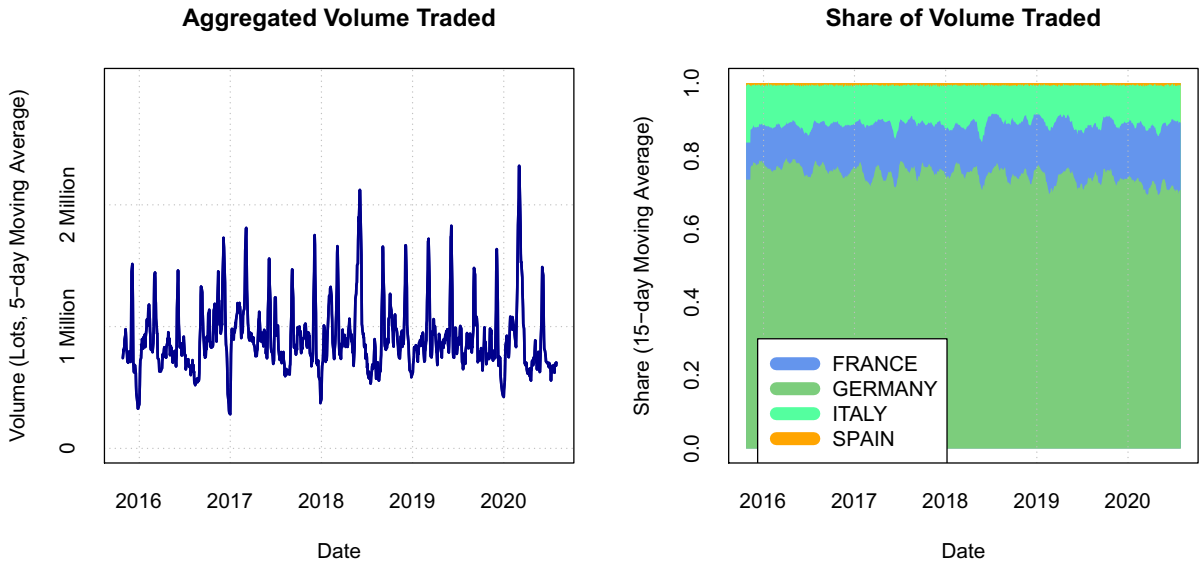


Figure 3.3: The left panel displays the 5-day moving average of the daily aggregated number of the most liquid maturity contracts traded for the French (OAT), German (BUND), Italian (BTP), and Spanish (BONO) government bond futures combined. The right panel displays the 15-day moving average of the aggregated daily volume proportion that can be attributed to each of the four countries' bond futures.

and considerable differences in the daily liquidity of the contracts considered in our study, with the German futures market unsurprisingly taking the decisive first place. Moreover, as the literature strongly suggests (McCauley et al. (1999), Ejsing and Sihvonen (2009) and Pelizzon et al. (2014)), the European sovereign bond futures market has been considerably more liquid for many years than the corresponding cash bond market. Because our study aims to document the return and volatility connectedness as well as the degree and direction of spillovers between the four major European sovereign bond markets, a comparison of the results with the futures market liquidity shares will also be of interest.

We transform the four non-stationary price series into daily log-return series by taking the first differences of the logarithms of daily closing prices. To avoid creating artificial "jumps" in daily log-returns on roll-over days, on each first day of the new maturity contract, we do not use the closing price of the previous maturity contract as the previous price. Instead, we use the closing price of the same contract, even though on previous day, that contract was not the most liquid one.

The summary descriptive statistics of the log-return series are displayed in Table 3.1. Panel A reports the statistics for the entire sample. We observe that, while all four securities recorded very similar average daily returns, the Italian BTP futures exhibit the greatest return variability, with a daily standard deviation of 0.55%. As Panels B and C suggest, considering the first

Table 3.1: Summary statistics of daily log-returns.

The table shows a selection of summary statistics for the daily log-returns of four long-term government bond futures considered in this study. While Panel A summarizes the entire sample period, Panels B and C display the statistics for two roughly equally divided subsamples. Mean, Median, Min, Max, and Standard Deviation statistics are expressed in percent. Kurtosis does not represent excess kurtosis. "Jarque Bera" shows the test statistics of the Jarque-Bera test of normality (null hypothesis: data is normally distributed). "ADF" displays the test statistic of the Augmented Dickey-Fuller test of covariance stationarity (null hypothesis: data is not covariance stationary). ACF(1) displays the lag-1 autocorrelation of log-returns. For the Jarque-Bera test, the ADF test and the ACF(1), the asterisk signs (*), (**) and (***) highlight the estimates that are statistically significant at the 10%, 5% and 1% level.

Panel A: Whole Sample

Time series:	FRANCE	GERMANY	ITALY	SPAIN
No. of Observations	1212	1212	1212	1212
Mean	0.02	0.02	0.02	0.02
Median	0.03	0.03	0.04	0.04
Min	-1.90	-1.84	-4.64	-2.41
Max	1.80	1.47	4.06	3.00
Standard Deviation	0.30	0.30	0.55	0.36
Skewness	-0.49	-0.44	-0.82	0.05
Kurtosis	4.29	2.85	11.62	10.79
Jarque-Bera	982(***)	451 (***)	6982 (***)	5906 (***)
ADF	-11.7 (***)	-12.2 (***)	-10.4 (***)	-11.5 (***)
ACF(1)	0.01	0.01	0.08 (***)	0.06 (**)

Panel B: October 26, 2015 - February 28, 2018

Time series:	FRANCE	GERMANY	ITALY	SPAIN
No. of Observations	599	599	599	599
Mean	0.01	0.01	0.01	0.01
Median	0.03	0.02	0.03	0.04
Min	-1.63	-1.48	-1.89	-1.80
Max	1.17	1.24	1.09	1.26
Standard Deviation	0.31	0.30	0.39	0.33
Skewness	-0.51	-0.42	-0.51	-0.47
Kurtosis	2.15	1.59	1.61	1.95
Jarque-Bera	143 (***)	81 (***)	91 (***)	118 (***)
ADF	-8.1 (***)	-8.3 (***)	-7.9 (**)	-8.8 (***)
ACF(1)	-0.05	-0.02	0.00	-0.01

Panel C: March 1, 2018 - July 31, 2020

Time series:	FRANCE	GERMANY	ITALY	SPAIN
No. of Observations	613	613	613	613
Mean	0.02	0.02	0.03	0.02
Median	0.04	0.04	0.04	0.04
Min	-1.90	-1.84	-4.64	-2.41
Max	1.80	1.47	4.06	3.00
Standard Deviation	0.29	0.30	0.66	0.38
Skewness	-0.45	-0.46	-0.82	0.37
Kurtosis	6.77	4.05	9.73	15.09
Jarque-Bera	1204 (***)	444 (***)	2507 (***)	5876 (***)
ADF	-10.2 (***)	-9.9 (***)	-8.3 (***)	-9.6 (***)
ACF(1)	0.07	0.03	0.10 (**)	0.11 (***)

and the second half of the sample, the increase in volatility of the BTP futures occurs in the second half of the sample, corresponding to the development of Italian political turbulence and the coronavirus outbreak. Moreover, with the exception of the Spanish BONO futures in the second half of the sample, all return series are negatively skewed (the BTPs to the greatest extent) and exhibit excess kurtosis. As the Jarque-Bera (JB) and the Augmented Dickey-Fuller (ADF) test results strongly suggest, all log-return series are non-normally distributed and are weakly stationary. Interestingly, during the second half of our sample, the Spanish and Italian futures log-returns are significantly positively first-order autocorrelated, with autocorrelation levels equal to approximately 10%.

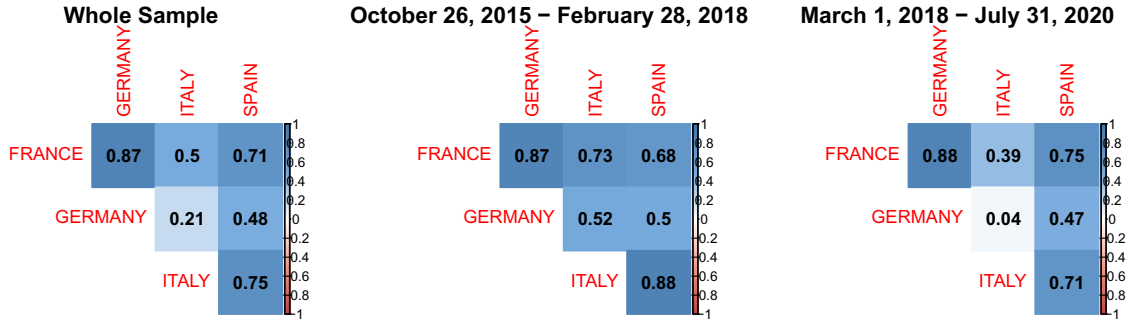


Figure 3.4: Cross-correlation coefficients of the four daily log-return series for the entire sample and for two roughly equally divided subsamples.

The cross-correlations of log-returns are displayed in Figure 3.4, again for the whole sample and for the two subsamples. The German and French bond futures are the most strongly correlated, with the coefficient equal to 87% for the entire sample, though not exhibiting much variability between the two subsamples. Italian bond futures are correlated with those of the other three countries to a lesser extent, especially in the second half of the sample. During this period, BUND and BTP futures are virtually uncorrelated, reflecting dynamic changes in the BTP-BUND yield spread occurring mostly due to the political developments in Italy.

We therefore continue our analysis of the stationary daily log-returns data. Apart from the return connectedness, we are also interested in investigating the spillovers of volatility. For each country, we take a non-parametric approach and estimate daily volatilities by calculating the rolling 30-day standard deviations of the log-returns. We then annualize these estimated daily standard deviations (assuming 252 trading days per year) and plot them in Figure 3.5.[13] Because the generalized forecast error variance decompositions assume normally distributed shocks, for robustness purposes, we also consider the logarithms of the volatility series, because, unlike the volatilities, the shape of the distribution of log-volatilities tends to be bell-shaped. We note that all 12 final time series (the four log-return series, the four volatility series, and the four

[13]By denoting the estimated daily standard deviation of log-returns as sd, the annualized values expressed in percet per annum plotted in Figure 3.5 correspond to $100 \cdot \sqrt{252 \cdot sd^2}$.

log-volatility series) are deemed to be weakly stationary, as strongly suggested by the ADF test. However, none of the series is deemed to be normally distributed according to the JB test, even though the log-volatility series reports smaller test statistic values than the volatility series.

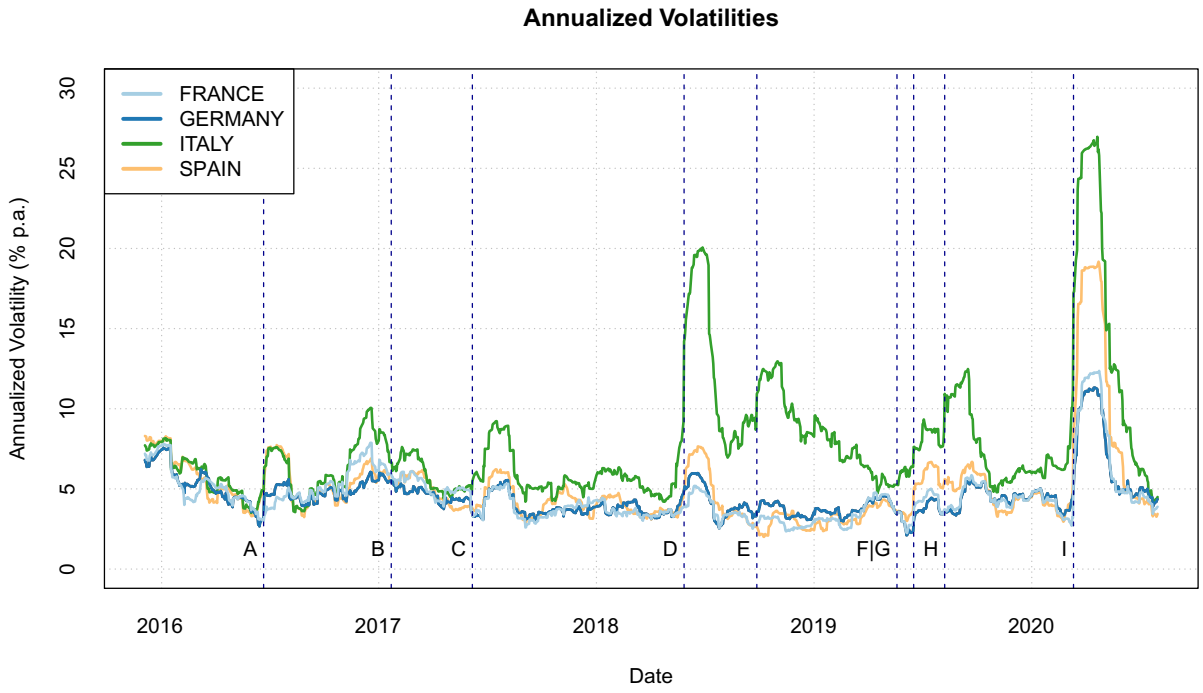


Figure 3.5: Estimated volatilities of the four daily log-return series. The volatilities are estimated as the 30-day rolling standard deviations of the daily log-return series. These estimated values are then annualized, assuming 252 trading days per year and expressed in percent per annum. The vertical labeled lines highlight selected political and macroeconomic events discussed in the main body of the text.

3.5 Results

3.5.1 The Static Whole Sample

We begin by estimating the VAR processes on the entire sample period and for two non-overlapping subsamples, setting March 1, 2018 as the division date. Table 3.2 displays the resulting 20-day forecast error horizon static connectedness matrices as well as the resulting static directional, net, and total spillover indices, considering all three types of input variables - log-returns, volatilities, and log-volatilities, as defined in 3.10, 3.11 and 3.12. We also report the lag orders of the estimated VAR processes as suggested by the combination of three information criteria. Further, we estimate the statistical significance of all connectedness measures using

bootstrapped standard errors. A vast majority of values is statistically significant at the 5% level, with the remaining insignificant values reported with an underscore.

We note that, in all three approaches, the Own Spillover Index explains the dominant part of the forecast error variance of each variable, as the diagonal terms of all three connectedness matrices are the maximum terms in each row. The greatest own-variable spillover share of 66.1% (62.8%) occurred in Italy's log-volatility during the second subsample (volatility during the entire sample). Conversely, French volatility appears to be the most affected by the volatility shocks of the other variables in the system, as only 28.3% of the forecast error variance is explained by shocks to French volatility itself across the entire subsample, particularly originating from the second subsample, when this reading changes to 23.9%.

Considering the spillovers in the log-return setting (Panel A of Table 3.2), we observe that Italian and German bond futures appear to be net log-return connectedness receivers, as opposed to Spanish and French bonds, for which we document positive net spillover values. For example, while the French OAT receives 60.3% of its error variance from the other variables, it contributes to the other variables by a total of 71.6% error variance, making it an 11.3% net transmitter of the log-returns error variance. We observe the largest pairwise spillover value of 37% from the French OAT to the German BUND forecast error variance, and document the smallest pairwise spillover of only 2.5% from German to Italian futures. Italy is also the country with the greatest own-variable spillovers, as 54% of the error variance in log-return forecasts is explained by shocks to the Italian log-returns themselves. Because Italy also reports the smallest directional "To" and "From" spillover indices, it appears to be the most isolated market in terms of log-return connectedness. The rationale for this isolation may be the turbulent national political developments that occurred during a significant period in our sample; particularly in the second half, as the second subsample spillover measures confirm.

We obtain dramatically different results, however, when we investigate the connectedness in volatility. As the volatility connectedness matrix reported in Panel B of Table 3.2 shows, the Italian BTP futures market appears to be the system's major volatility transmitter, with a whole sample net volatility spillover measure equal to 50.3%. Indeed, while the BTP receives 37.2% of the forecast error variance from others, it contributes by as much as 87.5% to the rest of the system in total. Albeit contrasting the log-return setting results in their magnitude, we propose the same rationale to the volatility results as well. Considering the fact that Italy's political crisis was a major factor present throughout our sample, and Eurozone membership of its third largest and second most indebted country was seriously in question during various stages of this political turmoil, our results suggest that the uncertainty in the form of greater forecast error variance was transmitted within the volatility setting to a considerably greater extent relative to the log-return setting. Therefore, the volatility of the bond futures of the other countries appears to be affected by Italian price dynamics more strongly than the actual levels

Table 3.2: Whole and (first half/second half) sample connectedness.

The table displays the estimated connectedness matrices C^H, as well as various spillover measures for the VAR processes estimated on the entire sample, starting on October 29, 2015 and ending on July 31, 2020 (top row for each country) and for two subsamples (October 29, 2015 - February 28, 2018 / March 1, 2018 - July 31, 2020, bottom row for each country). The forecast error horizon is $H = 20$ days. The information criteria - suggested lag order values are displayed for each VAR estimation. Panels A, B and C employ variables $Y^{(1)}_{i,d}, Y^{(2)}_{i,d}, Y^{(3)}_{i,d}$, defined in equations 3.10, 3.11 and 3.12. See section 3.3 for the definitions of all measures reported. The statistical significance of the values reported in the table is assessed using the non-parametric bootstrap methodology (described in section 3.3) with 3,000 repetitions. Values deemed insignificant at the 5% level are underlined. For the sake of brevity, we do not report the standard errors in this table (see Table 3.4 in Appendix 3.A for the standard errors of estimations for the entire sample).

Panel A: Static Log-Return Connectedness

	FRANCE	GERMANY	ITALY	SPAIN	Dir. From Others $\rho^H_{i\leftarrow\cdot}$
FRANCE	39.7	30.2	9.9	20.1	60.3
	(36.4/40.7)	(27.4/31)	(19.2/6.5)	(17/21.8)	(63.6/59.3)
GERMANY	37.0	48.6	2.7	11.7	51.4
	(33.2/38.6)	(44/48.6)	(11.7/1.8)	(11.1/11)	(56/51.4)
ITALY	13.0	2.5	54.0	30.5	46.0
	(20.5/8.4)	(10.3/0.5)	(38.9/60)	(30.3/31.1)	(61.1/40)
SPAIN	21.6	10.0	24.9	43.5	56.5
	(18.7/23.1)	(10.2/8.9)	(31.2/23.7)	(40/44.3)	(60/55.7)
Dir. To Others $\rho^H_{i\rightarrow\cdot}$	71.6	42.7	37.5	62.3	**TSI ρ^H_{TOT}: 53.6**
	(72.4/70.1)	(47.9/40.3)	(62.1/32)	(58.4/63.9)	**(60.2/51.6)**
Net Spillovers ρ^H_i	11.3	-8.7	-8.5	5.8	Lag Order: 4
	(8.8/10.7)	(-8.1/-11.0)	(1.0/-7.9)	(-1.7/8.2)	(1/4)

Panel B: Static Volatility Connectedness

	FRANCE	GERMANY	ITALY	SPAIN	Dir. From Others $\rho^H_{i\leftarrow\cdot}$
FRANCE	28.3	20.5	24.5	26.6	71.7
	(38.6/23.9)	(26.9/21.1)	(17.8/22.7)	(16.7/32.3)	(61.4/76.1)
GERMANY	20.8	29.4	26.0	23.9	70.6
	(24.1/20.5)	(32.5/30.4)	(21.2/23.5)	(22.2/25.7)	(67.5/69.6)
ITALY	6.1	9.2	62.8	21.8	37.2
	(14.1/3.4)	(11.0/14.5)	(42.8/62.7)	(32.1/19.4)	(57.2/37.3)
SPAIN	11.6	11.8	37.0	39.5	60.5
	(9.7/13.5)	(11.1/14)	(32.3/34.5)	(47/38.1)	(53/61.9)
Dir. To Others $\rho^H_{i\rightarrow\cdot}$	38.5	41.6	87.5	72.4	**TSI ρ^H_{TOT}: 60.0**
	(47.9/37.4)	(49/49.6)	(71.3/80.6)	(70.9/77.4)	**(59.8/61.2)**
Net Spillovers ρ^H_i	-33.2	-29.0	50.3	11.9	Lag Order: 8
	(-13.5/-38.7)	(-18.5/-20.1)	(14.0/43.3)	(17.9/15.5)	(3/8)

Panel C: Static Log-Volatility Connectedness

	FRANCE	GERMANY	ITALY	SPAIN	Dir. From Others $\rho^H_{i\leftarrow\cdot}$
FRANCE	35.2	24.5	15.6	24.8	64.8
	(40.0/30.5)	(27.2/22.8)	(14.8/16)	(18.1/30.7)	(60.0/69.5)
GERMANY	25.7	36.3	16.6	21.4	63.7
	(25.2/25.4)	(32.5/35.1)	(18.9/16.1)	(23.4/23.5)	(67.5/64.9)
ITALY	7.0	10.1	58.8	24.0	41.2
	(11.1/3.2)	(7.7/10.8)	(47.1/66.1)	(34.1/19.9)	(52.9/33.9)
SPAIN	12.7	12.0	27.6	47.6	52.4
	(9.4/13.8)	(7.4/11.9)	(31.7/27.9)	(51.5/46.4)	(48.5/53.6)
Dir. To Others $\rho^H_{i\rightarrow\cdot}$	45.4	46.6	59.8	70.2	**TSI ρ^H_{TOT}: 55.5**
	(45.7/42.4)	(42.3/45.5)	(65.3/60.0)	(75.6/74.1)	**(57.2/55.5)**
Net Spillovers ρ^H_i	-19.4	-17.0	18.6	17.8	Lag Order: 5
	(-14.4/-27.1)	(-25.2/-19.4)	(12.4/26.0)	(27.2/20.4)	(2/6)

of log-returns.

Moreover, considering the relatively small overall share of volume traded in the Italian BTPs (discussed in section 3.4), documenting such large and strong net spillover values suggests one possible way in which our connectedness measures may provide additional useful information regarding the system's risk contagion structure. Indeed, despite being the most liquid assets in our sample, both the German and the French futures act as overall net volatility receivers, as their net spillovers measures (equal to -29% and -33.2%, respectively, for the whole sample and appearing to be robust between the two subsamples) strongly suggest. Among the pairwise volatility spillovers, the largest value of 37% can be observed in the "from Italy to Spain" direction, while the smallest pairwise spillover of 6.1% in the "from France to Italy" direction (which declines into statistically insignificant territory during the second subsample) underlines Italy's major volatility transmitter status. French OAT futures contribute to the rest of the system with the smallest share of the forecast error variance at 38.5%. Overall, we observe considerably greater absolute values of the net volatility spillover indices than in the log-return setting. However, we note that the bootstrapped standard errors of these indices are greater as well, rendering the Spanish whole sample Net Spillover Index value positive, but statistically insignificant at the 5% level.[14]

For robustness purposes, we also report the log-volatility connectedness results in Panel C. Overall, though the results are qualitatively comparable to the spillovers in volatility, we observe some differences. For instance, although the net log-volatility spillovers maintain the same ordering among the variables, the absolute values of the spillovers are approximately 50% smaller. Nevertheless, Italy still acts as the major log-volatility transmitter, though the Spanish BONO reports only marginally smaller, though statistically insignificant, positive net spillover index. German and French bond futures remain the net log-volatility receivers, as the fraction of the forecast error variance they contribute to the Spanish and Italian futures is about two thirds of what is contributed in the opposite direction. The largest and smallest pairwise spillovers are observed for the identical pairs of variables as in the volatility setting.

We also note that, for all three settings, the value of the Total Spillover Index ranges from 53.6% to 61.2%, does not differ much between the two subsamples, and is greater for the two volatility settings, suggesting that the Eurozone bond futures market as represented by our sample constitutes a considerably interconnected system, and the possibility of risk contagion needs to be taken into account by both investors and regulatory institutions. However, when contrasted with the 78.3% total volatility connectedness of US financial institutions (as estimated by Diebold and Yılmaz (2014)), or with the extreme 97.2% volatility connectedness of international financial markets (Diebold and Yilmaz (2015)), we state that the degree of finan-

[14]See Table 3.4 in Appendix 3.A for the standard errors.

cial contagion in European major sovereign bond markets is moderate.[15] However, apart from the static results, it is of substantial interest to investigate the dynamic development of these spillover measures over time more closely. It is that form of analysis with which we now proceed.

3.5.2 The Dynamics of Connectedness

Total Spillover Index

We apply the VAR estimation to the 252-day rolling window subsamples to investigate the dynamics of the connectedness over time. For each subsample, this estimation produces the entire table of results, analogous to the results reported in Table 3.2. However, since displaying the entire connectedness matrix for every estimation (and for all three settings) is intractable, in this section, we focus on the time development of the following key spillover indicators: the Total Spillover Index, the country-wise "Own" Spillover Indices, and the directional "To", "From" and Net Spillover Indices.

Firstly, we turn our attention to the dynamics of the Total Spillover Index (ρ^H_{TOT}). Figure 3.6 displays the evolution of this index for each of the three settings considered in this study, depending on the variable modelled by the VAR process. Even though all three variables (the log-return levels, their volatilities, and log-volatilities) share certain features of the overall dynamics of total spillovers, noticeable differences can be seen.

Until May, 2018, very similar dynamics for all three types of variables were observable. The Total Spillover Index of all three systems oscillated around the level of 60%, recording two spikes in the second half of January, 2017 (B) and in June, 2017 (C). In the first case, the largest increase in the log-return (volatility) connectedness of 1.59 percentage points (p.p.) (1.64 p.p.) occured between January 23 and 24, 2017. On January 23, President Trump's remarks on the substantial planned tax cuts in the United States resulted in a surge in volatility across asset classes worldwide.[16]. During the upcoming days, our sample's connectedness continued to increase slightly, resulting in an approximately 5 p.p. increase overall in a one week period for all three classes of variables. A similar spike in connectedness measures occured in mid-June. In particular, on June 8, as the Italian electoral reform deal unravelled, the Total Spillover Index of volatility increased by 4 p.p. [17]

Subsequently, while the total volatility spillovers continued to increase until September, 2017, the log-return spillovers stalled. All three classes of variables then exhibited similar and relatively stable total connectedness until the end of May, 2018 (D), when a rapid selloff of

[15] Fernández-Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2016) report a value of 54.23% for the volatility connectedness of 11 European sovereign bond markets, as estimated on daily 10-year bond yields during the period 1999 - 2014.

[16] `https://www.bloomberg.com/news/videos/2017-01-23/trump-we-will-cut-taxes-massively`

[17] `https://www.reuters.com/article/us-italy-politics-idUSKBN18Z1JE`

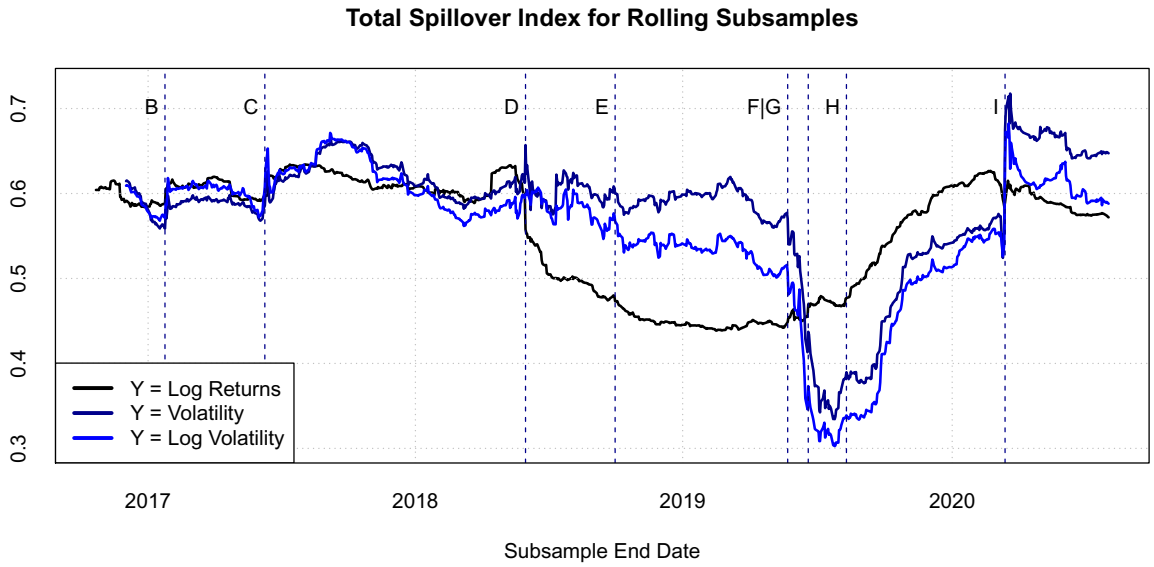


Figure 3.6: The time development of the Total Spillover Index ρ^H_{TOT} estimated on rolling subsample windows for the forecast error horizon of 20 days. The length of one subsample is 252 days, thus, the subsample start date is approximately one year earlier than the subsample end date (displayed on the x-axis). We plot all three classes of variables entering the underlying VAR processes: the log-return levels, their volatility, and their log-volatility. We chose the lag order of each of the VAR process estimations according to the procedure described in section 3.3. The vertical labeled lines highlight selected political and macroeconomic events discussed in the main body of the text.

Italian BTPs (accompanied to a lesser extent by Spanish BONO futures) resulted in a surge of volatility across all four assets, which continued for the next several weeks. However, while the volatility of the French, German, and Spanish futures returned to their pre-May levels by July, 2018 (as shown in Figure 3.5), the volatility of Italian futures maintained levels more than double of those of the other countries until March, 2019. This period of Italian bond market turbulence is accompanied by intriguing dynamics in the Total Spillover Indices. The most apparent feature is the divergence of the log-return and volatilities settings. Indeed, the total connectedness in log-returns exhibited a gradual, but substantial decrease of almost 20 p.p., declining to a minimum of about 44% by March, 2019. However, by the same date, the volatility connectedness maintained its long-term levels of around 60%, despite recording two rapid daily increases of about 4 p.p. (on May 29 and July 9), that were corrected within several days.

Interestingly, during the second quarter of 2019, the dynamics in the trends of the Total Spillover Indices changed markedly. The log-return connectedness gradually, but persistently, bounced upwards from its March minimum levels, reaching 61% by the end of the year. Meanwhile, the connectedness of both volatility measures experienced a pronounced U-shaped trajec-

tory. Firstly, during the period accompanied by increasing Brexit deal uncertainty (F) and the ECB's monetary policy easing signals (G), both volatility spillover indices declined by more than 40% (or approximately 20 p.p.) over the course of several weeks, reaching the overall minimum levels of 33.5% (30.3%) on July 24, 2019 (shortly before another volatile political development in Italy, starting with the successful no-confidence vote on August 9 (H)) and dropping below the log-return connectedness a few weeks earlier. Subsequently, their trends briskly reversed and the indices started climbing in a gradual, less volatile, but substantial manner. The upward trend of all three indices continued until the end of February, 2020, when the log-returns (volatility) connectedness measures reached the levels of 63% (56%).

On March 12, 2020 (I), the greatest daily increase in volatility, as well as log-volatility connectedness of 12.1 (9.7) p.p. was recorded, as fears of a severe global impact of the coronavirus Covid-19 outbreak continued to manifest in global financial markets (the S&P 500 Index, for example, plunged 9.5 percent on this day, its steepest one-day fall since 1987). The log-return connectedness, however, increased by only 0.6 p.p. on this day. Overall, over the course of one week, the log-return, volatility and log-volatility Total Spillover Indices rose by 2, 16, and 15 percentage points, respectively, marking an unprecedented increase in the volatility connectedness of the entire system. By the end of July, 2020, all three spillover indices had declined moderately (albeit remaining at historically large levels), as the monetary and the fiscal rescue packages in all major economies managed to tame the considerable anxiety prevailing in financial markets worldwide.

Having commented on their time development, we state that our results for the Total Spillover Indices vary considerably over time, corresponding to regional as well as global turbulence, with a tendency to increase when unanticipated news with a significant price impact (notably B, C, D, and I) arrives. Moreover, as the period 2018 - 2019 suggests, the log-return and the volatility spillover indices are not necessarily correlated and may diverge due to being driven by different market characteristics. Finally, the volatility and the log-volatility connectedness are correlated to a significant degree and do not exhibit substantial qualitative differences. Because of this observation and for the sake of brevity, for the remainder of our analysis, we will focus only on the log-return and the volatility connectedness settings.[18]

Analyzing the Total Spillover Index is appealing, since the entire complexity of a connectedness matrix is conveniently summarized in one number. However, important features of the nature and origins of connectedness may remain unnoticed. Therefore, we now turn to investigating more closely the dynamics of the country-wise Directional "From", "To", as well as the Net and Own spillover indices.

[18]The remaining results for the log-volatility setting are reported in Appendix 3.A.

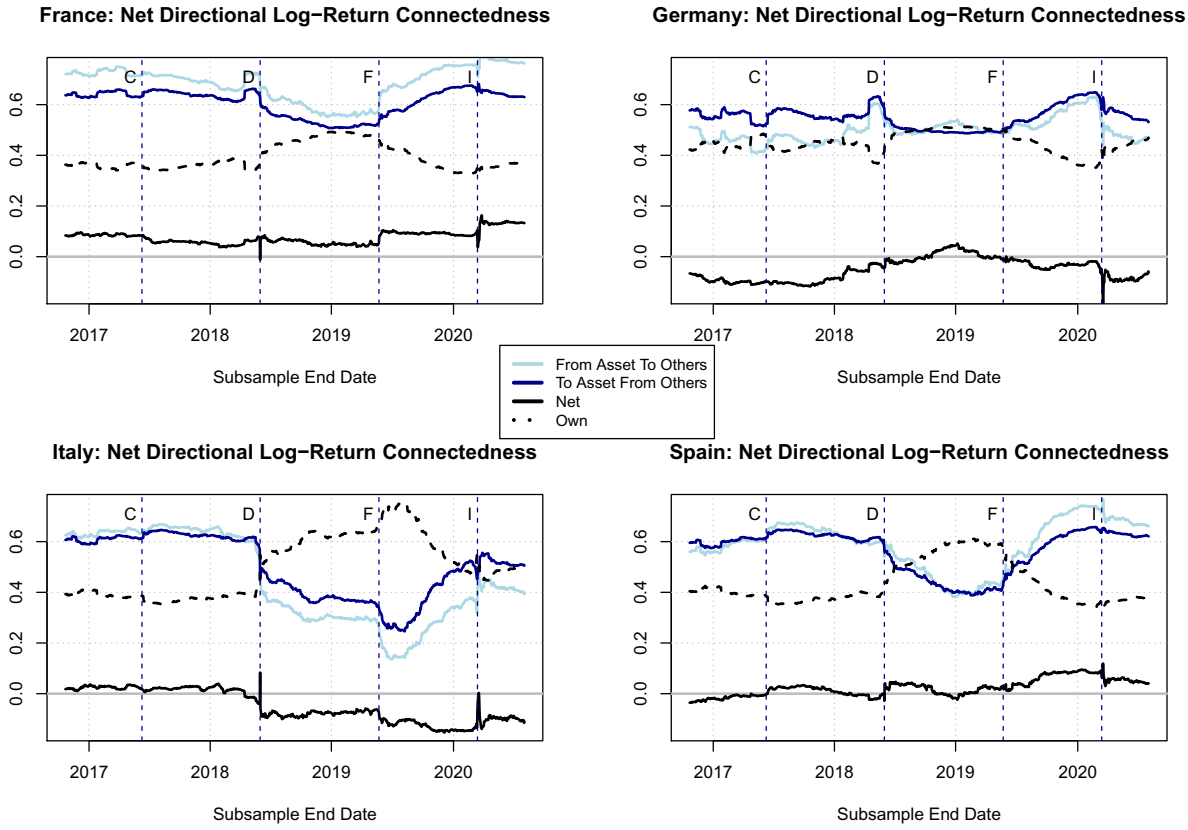


Figure 3.7: Development over time of the Directional Spillover Index "From" $\rho^H_{i\leftarrow\cdot}$, "To" $\rho^H_{i\rightarrow\cdot}$, the Net Spillover Index ρ^H_i, and the Own Spillover Index c^H_{ii} estimated for each of the four securities on rolling subsample windows for the forecast error horizon of 20 days. The variables entering the VAR estimations are the securities' log-returns. The length of one subsample is 252 days, thus, the subsample start date is approximately one year earlier than the subsample end date (displayed on the x-axis). We chose the order of each of the VAR process estimations according to the procedure described in section 3.3. The positive (negative) values of the Net Spillover Index correspond to variable i being a net transmitter (net receiver) of spillovers within the system. The vertical labeled lines highlight selected political and macroeconomic events discussed in the main body of the text.

Figure 3.7 and Figure 3.8 display the time development of the two directional spillover indices ("From" and "To"), their difference (the Net Spillover Index), as well as the Own Spillover Index for all four countries. An asset with a positive value of the Net Spillover Index can be regarded as the net spillover transmitter, because it contributes more overall forecast error variance shares to the system than it receives from the system. Conversely, an asset with a negative Net Spillover Index acts as the net spillover receiver. Figure 3.7 considers the log-return spillovers setting, while Figure 3.8 displays the results for the volatility spillovers VAR estimations.

Considering first the log-return setting shown in Figure 3.7, we state that, in line with the

whole sample results reported in Table 3.2, France appears to be the principal net log-return connectedness transmitter, as its value of the net spillovers remained positive throughout our sample, increasing in a step-wise fashion following the May 2018 Italian political (D) and May 2019 (F) Brexit deal turbuleces, and the March 2020 coronavirus stress period (I). Similar behaviour of net spillovers, albeit to a lesser extent, are evident for Spain, as is also suggested in Table 3.2. Conversely, until May 2018, and despite its dominant liquidity position (in terms of overall trading volume), Germany consistently recorded net spillover levels of around -10%. Germany's role as the main net receiver of log-return spillovers was, however, taken over by Italy, following the May 2018 turbulence (D). Interesting dynamics may be observed especially during the two system-wide shocks discussed above - the BTP and BONO selloff on May 29, 2018 (D) and the coronavirus outbreak in mid-March 2020 (I).

During the first shock, on May 29, 2018, Italy's net return connectedness jumped from -3% to about 9%, immediately reversing the next day to a sub-zero level of -6% and then remaining negative. Similarly, Span's net return connectedness increased by approximately 6 p.p., while the French Net Spillover Index recorded a negative one-day spike of about 8 p.p.. We also note that the events of May 29 can be considered to have been the main triggers of greater isolation of the log-return dynamics for all four countries, as the increasing trajectories of the Own Spillover Indices starting on this day (and lasting for almost two years) document. The greatest isolation may be decidedly observed for the "periphery" countries, particularly Italy (which almost doubled its Own Spillover Index, peaking in mid-2019), and Spain (which experienced an approximately 50% increase of its own spillovers). Overall, this period of greater isolation unsurprisingly corresponds to the wide U-shaped trajectory of the total log-return connectedness discussed earlier. As these net spillover data suggest, even though Italy isolated to the greatest extent, its Directional Index "From" declined more than the Directional Index "To". Hence, after May, 2018, Italy became the strongest net receiver of log-return forecast error variance. We note that, in mid-2019, the overall system-wide isolation began to weaken, as the own spillovers of all four countries gradually decreased, resulting in the system's greater total log-return connectedness.

The other major shock to the system, starting on March 12, 2020 and lasting for about 4 business days, took place in the midst of the first European wave of Covid-19. While French and German net log-return connectedness spiked downwards during this period, the trajectory was reversed for Italy and Spain. The magnitude of spikes was the greatest for Germany (-15 p.p.) and Italy (+13.5 p.p.). Again, these rapid changes in connectedness were corrected within several business days. Nevertheless, this episode highlights the already observed differentiation of our sample into two groups of countries, namely, the "central" countries of Germany and France, and the "periphery" countries of Italy and Spain.

It is interesting to contrast these observations with the dynamics of volatility spillovers,

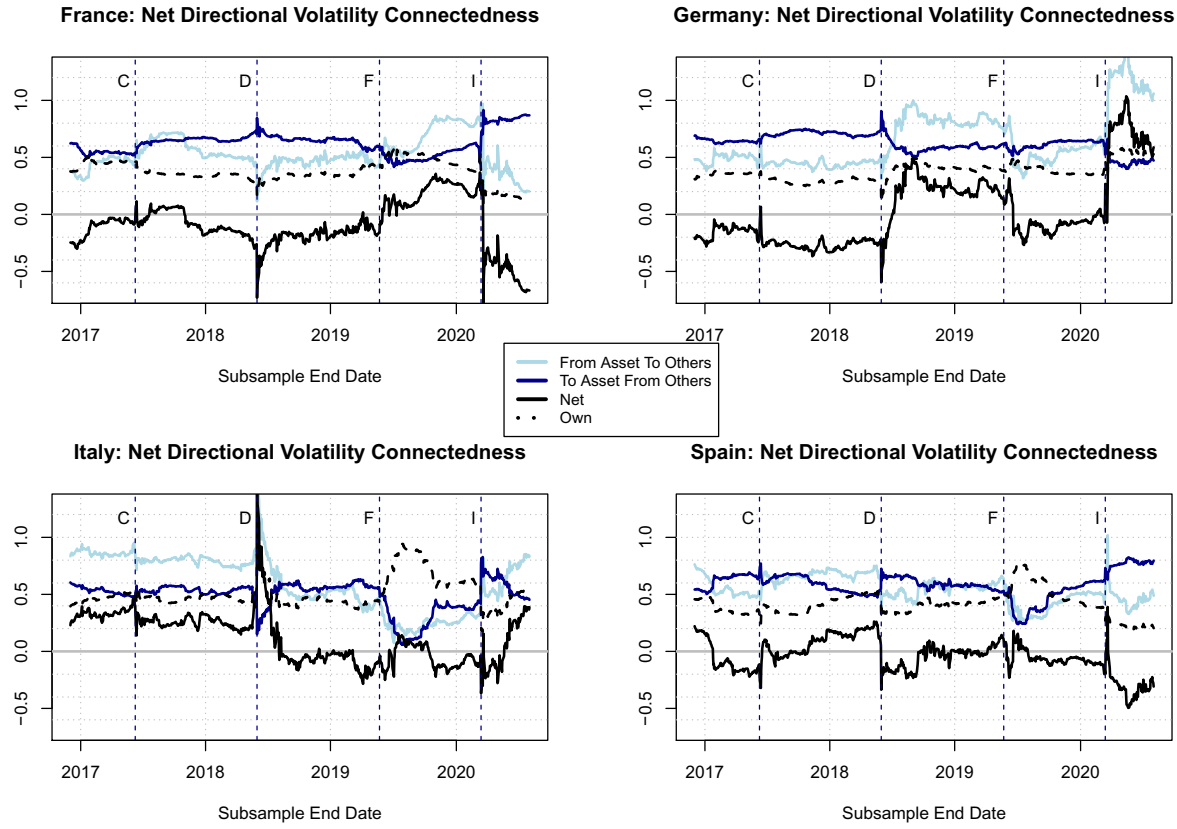


Figure 3.8: Development over time of the Directional Spillover Index "From" $\rho^H_{i\leftarrow\cdot}$, "To" $\rho^H_{i\rightarrow\cdot}$, the Net Spillover Index ρ^H_i, and the Own Spillover Index c^H_{ii} estimated for each of the four securities on rolling subsample windows for the forecast error horizon of 20 days. The variables entering the VAR estimations are the securities' volatilities. The length of one subsample is 252 days, thus, the subsample start date is approximately one year earlier than the subsample end date (displayed on the x-axis). We chose the order of each of the VAR process estimations according to the procedure described in section 3.3. The positive (negative) values of the Net Spillover Index correspond to variable i being a net transmitter (net receiver) of spillovers within the system. The vertical labeled lines highlight selected political and macroeconomic events discussed in the main body of the text.

shown in Figure 3.8. Due to the noisier estimates and the overall greater standard errors for the volatility setting (also documented in Appendix 3.A), we focus primarily on the major peaks, troughs, and sudden changes in the Net Spillover Indices. Therefore, we reiterate the importance of the following major shocks observed in our sample: the Italian political events on June 8, 2017 (C) and May 29, 2018 (D), the mid-2019 mixture of Brexit-, ECB-, and Italy-related developments (F) and the coronavirus outbreak volatility shock ignited on March 12, 2020 (I).

The Italian net volatility spillover index recorded values in the range of 15% - 30% for the majority of days until May 15, 2018 (rendering Italy the main volatility spillovers transmitter during this period, while Germany served as the main volatility spillovers receiver), holding a historically unremarkable value of 20% on this day. Subsequently, the value of this index

more than doubled over the next four days, and then maintained levels around 42% until May 28. On May 29 (D), the Italian Net Spillover Index soared to an unprecedented level of 166% and remained above 100% for the next five days, while the Italian BTP market transmitted a remarkable amount of uncertainty to the system and the other three countries all became major receivers of volatility spillovers, with their Net Spillover Indices recording substantial negative jumps at magnitudes of approximately 40 percentage points. Following these events, however, Italy's Net Spillover Index dropped to statistically insignificant territory within several weeks, while Germany gradually took over the role of the main transmitter of volatility until mid-2019.

Similarly to the trajectory of the log-returns setting, the Total Spillover Index in the volatility setting also recorded a pronounced "isolationist" U-shaped path, though it occurred in a different time period, starting a downward trend in the beginning of 2019, reaching a trough in mid-2019, and returning to original levels by the end of 2019. We state that both U-shaped trajectories are driven mainly by the corresponding inversely U-shaped dynamics of the individual Own Spillover Indices. Among these, the Italian and Spanish Own Spillover Indices again play dominant roles. Therefore, unlike the log-return setting isolation that started immediately after the shock of May 29, 2018, in the case of the volatility setting, this period of heightened isolation took place in its entirety during 2019.

Regarding the second interesting spike episode that occurred during the first wave of Covid-19-related market distress, German net volatility spillovers jumped from 3.5% to 24% on March 12, 2020 (I). Subsequently, the index rapidly increased to about 80% and reached a record high of 103% on May 12, 2020. As of July 31 (the end of our data sample), the German net volatility spillover index equalled approximately 58%, still rendering Germany the main volatility transmitter, followed by Italy. Conversely, France and Spain were the system's net receivers of volatility spillovers during the same period.

Based on our analysis, we state that prior to, and especially during the shock of May, 2018, Italian internal political affairs dramatically affected not only the prices of Italian government bonds, but significantly contributed to the forecast error variances of both the log-returns levels and the volatilities of the other three EU sovereign bond markets. Subsequently, until the second shock, the majority of volatility spillovers was transmitted either by Germany or by France, and received by the other markets. Because this period corresponds to the development of multitude of global geopolitical and economic affairs (including US-China trade tarrifs, Brexit deal negotiations and the monetary policy-easing rhetoric of the FED and the ECB), the bond markets of the "central" countries of Germany and France can be considered to be the channels through which these events contributed to the price discovery in the "periphery" EU sovereign bond markets of Italy and Spain. In this respect, we assert that Germany takes the leading role of the sample's volatility transmitter of global affairs, as the second shock, corresponding to the coronavirus outbreak and its development, strongly documents.

We therefore conclude that log-return and volatility spillovers settings do not neccessarily produce equivalent results in terms of the origins, magnitudes, and directions of system-wide connectedness. Indeed, while individual net spillovers in the log-returns setting never exceeded the level of 20% in their absolute value, it was not uncommon to observe levels beyond this threshold for the net volatility spillovers setting, especially during the more turbulent periods. Moreover, we note that opposite characteristics of spillovers directions can be observed during crisis periods (such as the initial coronavirus shock period, when Germany and Italy received most of the log-return spillovers and simultaneously transmitted most of the volatility spillovers, while the opposite behaviour was exhibited by France and Spain). Additionally, as illustrated on the two major shocks present in our sample, the dynamics of these net directional characteristics are especially interesting and may deserve particular attention during periods of such significant market distress. Therefore, an investigation of the possible association between these dynamics and fundamental macroeconomic and financial indicators follows.

3.5.3 Determinants of Directional Spillovers

The association between underlying macroeconomic conditions and European sovereign bond market characteristics presents a popular focus of academic studies; for example, see Balli (2009), Favero and Missale (2012), Costantini, Fragetta, and Melina (2014), Gómez-Puig, Sosvilla-Rivero, and del Carmen Ramos-Herrera (2014), Afonso et al. (2015), and Fernández-Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2016). We join this stream of literature by investigating the potential association of bond market connectedness calculated in this study (within our set of four EU countries) and the variables capturing the contemporary economic conditions. Considering the sample size at our disposal and following this stream of recent relevant studies,[19] for each country and each calendar month, we collect the following set of macroeconomic and market sentiment variables:

- Real GDP growth ($GDPGrowth$),
- General government consolidated gross debt as a percentage of GDP ($DebtToGDP$),
- Harmonized Index of Consumer Prices (CPI),
- Unemployment rate as a percentage of active population ($Unem$),
- Consumer confidence indicator ($ConsConf$),

[19] Especially Fernández-Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2016).

- The standard deviation of daily log-returns of the corresponding country's stock market index during the corresponding calendar month ($MarketSD$).[20]

We source the macroeconomic variables from Eurostat and the daily stock market data from Bloomberg. We use seasonally adjusted data, whenever applicable. For $GDPGrowth$ and $DebtToGDP$, which are sourced at quarterly frequency, we use linear interpolation for the within-quarter months to obtain monthly data. Additionally, based on exploratory analysis, we transform selected variables as follows: we take the first differences of the $DebtToGDP$ and $Unem$ variables to account for the strong persistence and the presence of unit roots (as confirmed by Augmented Dickey-Fuller tests) and thereby create variables $D(DebtToGDP)$ and $D(Unem)$. Finally, due to its high skewness and asymmetric distribution, we take the logarithms of the $MarketSD$ variable and create the variable $Log(MarketSD)$.

Since a comparison of the spillover dynamics between more turbulent and more tranquil episodes is of considerable interest, we additionally create a monthly dummy variable $Crisis$, equal to 1 if the particular calendar month belongs to a more turbulent episode and 0 otherwise. Based on our earlier estimation results, and combined with geopolitical and macroeconomic developments, we consider two disjointed periods to be more turbulent periods: April, 2018 until October, 2018 and January, 2020 until July, 2020 (the end of our sample). The first "crisis" period contains the turbulent Italian political developments, including the rapid BTP selloff culminating on May, 29, and the onset of the potential US-China trade war dispute and the Brexit deal negotiations. The second period contains the worldwide Covid-19 outbreak.

Regarding the dependent variables, for each country and each calendar month, we calculate the arithmetic averages of the daily Net Spillover Indices obtained in the preceding rolling subsample analysis (and displayed in Figure 3.7 and Figure 3.8), both for the log-return setting ($NetSpilloverLogR$) and for the volatility setting ($NetSpilloverVol$). Since we aim to investigate the possible macroeconomic determinants of these net spillover indices, we use panel data regressions, allowing for both time-specific and country-specific fixed effects.[21] To account for the possibly complicated (and fully general) clustering structure of the residual covariance matrix, we assess the statistical significance of the estimates using White heteroskedasticity-robust standard errors as proposed by Arellano et al. (1987).

The regression results for both types of dependent variables are reported in Table 3.3. For each dependent variable, we follow the general-to-specific approach (introduced by Hendry et al. (1995) and described in detail by Faust and Whiteman (1997)). We start with the most general

[20]We choose the following stock market indices: CAC40 for France, DAX for Germany, FTSEMIB for Italy and IBEX for Spain. The daily data are sourced from Bloomberg, tickers: CAC Index, DAX Index, FTSEMIB Index, IBEX Index.

[21]We apply the F-test to test for the presence of fixed effects in both dimensions to assess whether the fixed effects or the pooled OLS approach is relevant in our setting. The results of the F-tests suggest the presence of fixed effects in both dimensions.

representation, containing all the regressors listed above and estimating the regressions both with and without the $Crisis$ interaction effects. We proceed by sequentially dropping the statistically insignificant variables and applying diagnostic checks to ensure the validity of model specification and residual properties. For the regressions that contain the $Crisis$ interaction terms, we drop the non-interaction terms only when both the non-interaction term and the corresponding interaction term are jointly insignificant, as identified by the F statistic.

Based on the reported results, in accordance with the observations made in the previous sections, that the log-return and volatility settings exhibit qualitatively different results. Considering firstly the "Without Dummy" regressions reported in columns (1) and (3), we note that the directional spillovers in log-returns and in volatility appear to be strongly associated with changes in the $DebtToGDP$ variable. However, while the estimated direction of the effect is positive for log-return spillovers, its direction in the volatility setting is reversed, such that a positive change in a country's indebtedness is associated with greater values of the log-return directional spillovers (making the country more likely to be a net transmitter of log-return spillovers) and smaller values of the directional spillovers of volatility (making the country more likely to be a net receiver of volatility spillovers). Moreover, for both types of spillovers, we document a relatively strong (at the 10% significance level) positive association with the corresponding country's stock market volatility. Therefore, we state that volatile stock market conditions accompany enhanced volatility in the corresponding country's sovereign bond market in a manner that also transmits spillovers into sovereign bond markets of the other countries. In the case of the volatility spillovers setting, we document an additional positive association between inflation and net volatility connectedness.

Moreover, dynamics of the association between directional spillovers and the economic fundamentals appears to differ between tranquil and turbulent periods, as the statistically significant interaction terms in regressions reported in columns (2) and (4) document. We state that the turbulent periods marked with the $Crisis$ attribute signal when this association dominates. In particular, we note that the directional differences between the log-return and volatility settings for the $D(DebtToGDP)$ remain, although they are now reflected in the opposite signs of the strongly statistically significant interaction term $D(DebtToGDP) * Crisis$, while the significance of the original slope variable coefficient disappears. This observation is in accordance with the connectedness dynamics we have discussed earlier. Another interesting effect is observed for the consumer confidence indicator: during non-crisis periods, its effect is slightly negative for the log-return connectedness and insignificant for the volatility connectedness, but during crisis periods, the observed association is positive (and statistically significant) for both types of spillovers.

Additionally, we observe negative and statistically significant interaction terms pertaining to the $GDPGrowth$ and CPI variables for the volatility connectedness regression. We assert that

Table 3.3: Determinants of the Net Spillover Index.

The table displays the results of the panel data regressions for both types of dependent variables ($NetSpilloverLogR$ and $NetSpilloverVol$) and for combinations of the explanatory variables that both include and exclude $Crisis$ dummy variable interaction terms. The $Crisis$ dummy variable is equal to 1 for periods of increased market distress (April 2018 - October 2018 and January 2020 - July 2020) and 0 otherwise. Please see the main text for more detailed description of the variables included. The data is sampled at monthly frequency. All models are estimated using time and country-wise fixed effects. The White heteroskedasticity-robust standard errors that account for possible complex clustering structure of the residual covariance matrix (due to Arellano et al. (1987)) are reported in parentheses. The asterisk signs (*), (**) and (***) highlight estimates that are statistically significant at the 10%, 5% and 1% level, respectively.

	Dependent variable:			
	NetSpilloverLogR		NetSpilloverVol	
	(1)	(2)	(3)	(4)
	Without Dummy	With Dummy	Without Dummy	With Dummy
D(DebtToGDP)	0.039***	0.018	−0.315***	−0.033
	(0.007)	(0.015)	(0.096)	(0.050)
D(DebtToGDP) * Crisis		0.049**		−0.492***
		(0.023)		(0.099)
GDPGrowth				0.132*
				(0.078)
GDPGrowth * Crisis				−0.119**
				(0.056)
CPI			0.082***	0.073
			(0.028)	(0.057)
CPI * Crisis				−0.135*
				(0.072)
ConsConf		−0.007**		0.002
		(0.003)		(0.011)
ConsConf * Crisis		0.006***		0.021***
		(0.002)		(0.007)
Log(MarketSD)	0.061*	0.071**	0.164*	
	(0.033)	(0.030)	(0.091)	
Log(MarketSD) * Crisis		−0.056*		
		(0.031)		
Observations	184	184	176	176
R^2	0.098	0.310	0.184	0.356
Adjusted R^2	−0.241	0.006	−0.133	0.084
F Statistic	7.222***	7.145***	9.498***	11.355***
Note:				*p<0.1; **p<0.05; ***p<0.01

during crisis periods, greater values of GDP growth and faster growth of the price level tend to decrease the (likely already present) levels of market distress, also decreasing the magnitude of the volatility spillovers a country's sovereign bond market is transmitting to the rest of the system. However, we acknowledge that, due to the negative sign of the $D(DebtToGDP)$ and the positive sign of the $ConsConf$ interaction terms, a similar reasoning does not apply to these two fundamental variables.

Nevertheless, we find some evidence that the documented relationships between macroeconomic and market sentiment variables vary between crisis and non-crisis periods. These observations are in line with the recent literature (for example, Gómez-Puig, Sosvilla-Rivero, and del Carmen Ramos-Herrera (2014), Afonso et al. (2015), and Fernández-Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2016)). In particular, the Fernández-Rodríguez, Gómez-Puig, and Sosvilla-Rivero (2016) study documents significant relationships between the connectedness of several European sovereign bond markets and underlying economy fundamentals. Moreover, these relationships vary between peripheral and central countries, and between crisis and non-crisis periods. Our findings are similar, and we therefore conclude that the dynamics of the cross-country risk contagion present in the markets analyzed in this study deserve careful attention from all relevant institutions. Particular attention is advisable during episodes of increased geopolitical, macroeconomic, or financial distress.

3.6 Conclusion

This study aims to complement the existing stream of empirical academic literature on spillovers in financial markets by carrying out a risk contagion analysis of the EU sovereign bond market during 2016 - 2020, a period including a wide range of macroeconomic and geopolitical events. In pursuit of this objective, we utilize the connectedness framework of Diebold and Yilmaz (2012) to study the developments over time and potential macroeconomic determinants of multiple directional and aggregate spillover measures on a set of the four most liquid European sovereign bond futures contracts (France, Germany, Italy, and Spain), focusing especially on the directions in which shocks to individual assets contribute to forecast error variances of the log-returns and volatility dynamics of the rest of the system.

We find that, although we observe a strong association between macroeconomic and geopolitical developments and spillover dynamics within the sovereign bond market, there are several dimensions in which the characteristics of the documented relationships differ substiantially. Firstly, the results differ with respect to the choice of the VAR estimation setting, depending on whether log-returns or volatilities are considered. Secondly, even within a particular setting (log-returns or volatilities), considerable differences between the spillover dynamics of the "central" countries of Germany and France and the "peripheral" countries of Italy and Spain

are not uncommon. While France appears to be the overall main transmitter of spillovers in the log-return setting and the main receiver of volatility spillovers, we observe the opposite characteristics for Italy, the sample's major volatility spillovers transmitter. Germany, despite being decisively the most liquid market in our study, appears to be a predominant receiver of log-return spillovers. However, we assert that, in various time periods, Germany may play the role of the major transmitter of volatility spillovers as strongly as that of their receiver. Interestingly, we note that Germany assumes the leading role as the sample's major volatility transmitter particularly during periods of increased global turbulence, as the periods including the late 2018 and early 2019 (the main US-China trade dispute and Brexit deal negotiations), and the most dramatic worldwide volatility shock in decades (the Covid-19 pandemic outbreak) strongly document. We conjecture that, while the volatility transmitting role of Italy was caused mainly by its internal turbulent political development, Germany takes the leading role as the sample's volatility transmitter of major global affairs.

Finally, as our panel data regression analysis suggests, the nature of the association of directional Net Spillover Index with underlying macroeconomic fundamendals appears to depend on the presence of crisis period as well as on the type of spillover considered. In particular, a change in a country's indebtedness appears to possess strong explanatory power, being positively associated with log-returns spillovers and negatively associated with volatility spillovers during crisis periods. A deeper investigation into the nature of this relationship is definitely worth pursuing and, together with extending the analysis by quantifying the spillovers using quantile-based measures, we suggest these as two possible directions for future research.

As we always report values of the Total Spillover Index below 72%, we deem the degree of risk contagion within the Eurozone sovereign bond market to be moderate with respect to the other benchmark studies. We find that all connectedness measures react rapidly to new information, as the spikes in their trajectories corresponding to major geopolitical shocks in our sample illustrate.

Our results appear to be robust with respect to our choices of hyperparameter values needed for the connectedness estimation procedure. Additionally, as the Italian sovereign bond market in our study demonstrates, we note that it is not unusual for a considerable amount of spillovers to be transmitted by a market whose liquidity, in terms of average traded volume, represents only a relatively small fraction of the system's aggregate liquidity. In this sense, we conclude that the connectedness measures investigated in this study can produce nuanced and elaborate insights into the structure of risk contagion dynamics potentially present in financial markets. Within the context of our analysis, we assert that the dynamics of cross-country risk contagion present in the European sovereign bond market deserve serious attention by fiscal and monetary policymakers, and of all market participants, particularly during episodes of increased geopolitical, macroeconomic, or financial distress.

3.A Appendix

In this section, we report the following additional results: Figure 3.9 shows developments over time of the Directional, Net, and Own Spillover Indices for the rolling subsample analysis that employs time series of log-volatilities as the main variables entering the VAR estimations. In this sense, Figure 3.9 complements Figure 3.7 (log-returns) and Figure 3.8 (volatilities). Table 3.4 complements Table 3.2 by containing standard errors based on the non-parametric bootstrap methodology using 3,000 bootstrap repetitions. With several exceptions (the underlined values), all estimated connectedness measures are statistically significant at the 5% level. Finally, Figure 3.10 displays the sensitivity of the estimated Total Spillover Index trajectory to the choice of the lag length of the uderlying VAR process (parameter p) and to the choice of the forecast horizon (parameter H). Overall, the resulting Total Spillover Index trajectories appear to be highly robust to the choices of these parameters.

Table 3.4: Whole sample connectedness with standard errors.

The table displays the estimated connectedness matrices C^H, and the defined spillover measures for the VAR processes estimated on the entire sample, starting on October 29, 2015 and ending on July 31, 2020. The forecast error horizon is $H = 20$. The information criteria - suggested lag order values are displayed for each VAR estimation. Panels A, B and C employ variables $Y_{i,d}^{(1)}, Y_{i,d}^{(2)}, Y_{i,d}^{(3)}$, defined in equations 3.10, 3.11 and 3.12. See section 3.3 for the definitions of these variables as well as of all the spillover measures reported. The standard errors are based on non-parametric bootstrap methodology (described in section 3.3) with 3,000 bootstrap repetitions.

Panel A: Static Log-Return Connectedness

	FRANCE	GERMANY	ITALY	SPAIN	Directional From Others $\rho_{i\leftarrow\cdot}^H$
FRANCE	39.7	30.2	9.9	20.1	60.3
Standard Error	1.24	1.45	1.38	0.98	1.24
GERMANY	37.0	48.6	2.7	11.7	51.4
Standard Error	1.20	1.84	0.89	1.50	1.84
ITALY	13.0	2.5	54.0	30.5	46.0
Standard Error	1.62	0.95	2.45	1.17	2.45
SPAIN	21.6	10.0	24.9	43.5	56.5
Standard Error	0.92	1.49	1.43	1.54	1.54
Directional To Others $\rho_{i\rightarrow\cdot}^H$	71.6	42.7	37.5	62.3	Total Spillover Index ρ_{TOT}^H: 53.6
Standard Error	1.78	2.67	2.80	1.96	1.52
Net Spillovers ρ_i^H	11.3	-8.7	-8.5	5.8	Lag Order: 4
Standard Error	0.82	1.33	1.16	0.92	

Panel B: Static Volatility Connectedness

	FRANCE	GERMANY	ITALY	SPAIN	Directional From Others $\rho_{i\leftarrow\cdot}^H$
FRANCE	28.3	20.5	24.5	26.6	71.7
Standard Error	2.48	3.52	3.32	2.67	2.48
GERMANY	20.8	29.4	26.0	23.9	70.6
Standard Error	2.62	3.57	3.56	2.72	3.57
ITALY	6.1	9.2	62.8	21.8	37.2
Standard Error	2.69	3.27	6.16	3.81	6.16
SPAIN	11.6	11.8	37.0	39.5	60.5
Standard Error	2.88	3.16	3.86	3.17	3.17
Directional To Others $\rho_{i\rightarrow\cdot}^H$	38.5	41.6	87.5	72.4	Total Spillover Index ρ_{TOT}^H: 60.0
Standard Error	7.49	9.04	9.95	8.27	1.90
Net Spillovers ρ_i^H	-33.2	-29.0	50.3	11.9	Lag Order: 8
Standard Error	9.34	12.08	14.06	10.53	

Panel C: Static Log-Volatility Connectedness

	FRANCE	GERMANY	ITALY	SPAIN	Directional From Others $\rho_{i\leftarrow\cdot}^H$
FRANCE	35.2	24.5	15.6	24.8	64.8
Standard Error	2.67	3.06	2.71	2.82	2.67
GERMANY	25.7	36.3	16.6	21.4	63.7
Standard Error	2.70	3.09	2.79	2.71	3.09
ITALY	7.0	10.1	58.8	24.0	41.2
Standard Error	2.40	2.94	4.96	3.08	4.96
SPAIN	12.7	12.0	27.6	47.6	52.4
Standard Error	2.80	2.82	3.07	4.01	4.01
Directional To Others $\rho_{i\rightarrow\cdot}^H$	45.4	46.6	59.8	70.2	Total Spillover Index ρ_{TOT}^H: 55.5
Standard Error	6.54	7.33	7.64	7.36	2.25
Net Spillovers ρ_i^H	-19.4	-17.0	18.6	17.8	Lag Order: 5
Standard Error	8.16	9.23	8.65	9.60	

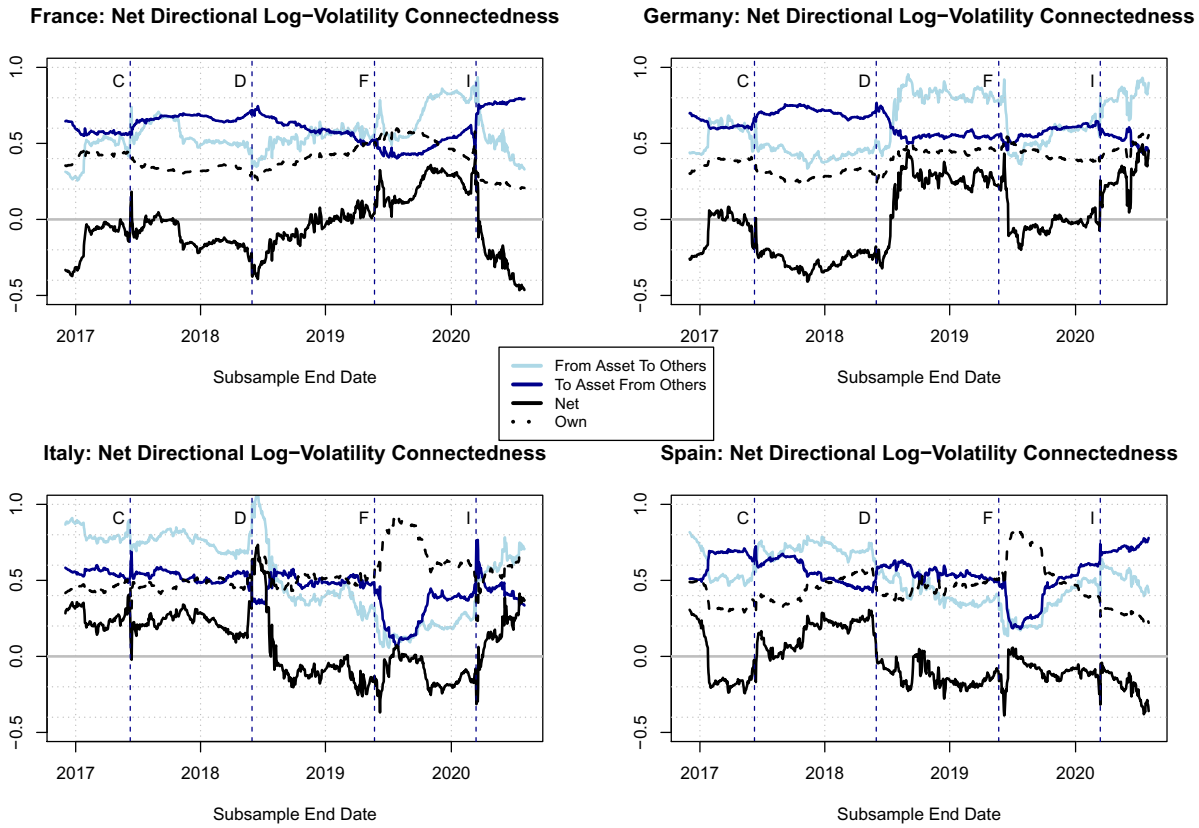


Figure 3.9: Development over time of the Directional Spillover Index "From" $\rho^H_{i\leftarrow\cdot}$, "To" $\rho^H_{i\rightarrow\cdot}$, the Net Spillover Index ρ^H_i, and the Own Spillover Index c^H_{ii} estimated for each of the four securities on rolling subsample windows for the forecast error horizon of 20 days. The variables entering the VAR estimations are the securities' log-volatilities. The length of one subsample is 252 days, thus, the subsample start date is approximately one year earlier than the subsample end date (displayed on the x-axis). We chose the order of each of the VAR process estimations according to the procedure described in section 3.3. The positive (negative) values of the Net Spillover Index correspond to variable i being a net transmitter (net receiver) of spillovers within the system. The vertical labeled lines highlight selected political and macroeconomic events discussed in the main body of the text.

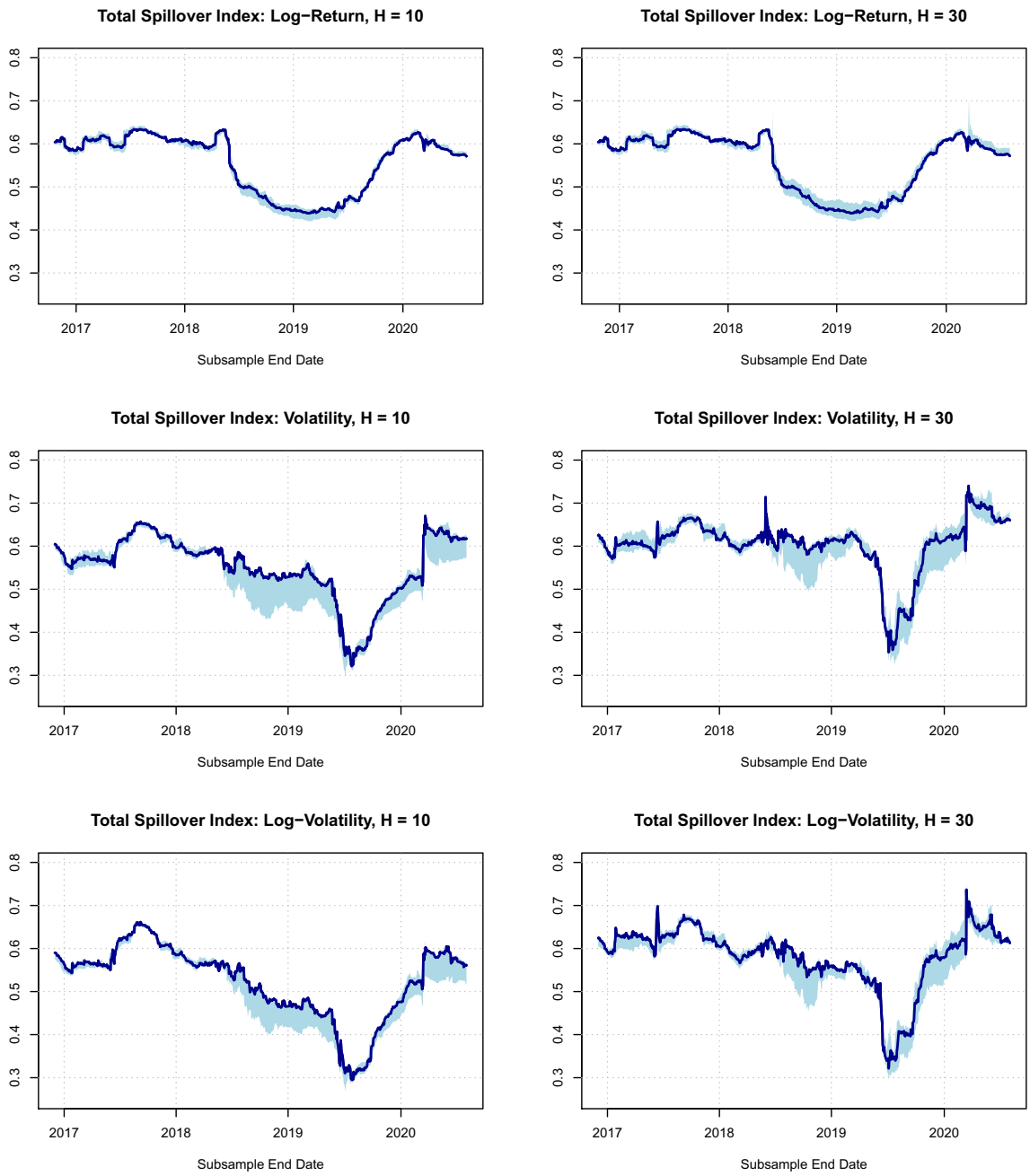


Figure 3.10: Robustness analysis: the developments over time of the Total Spillover Index ρ^H_{TOT} estimated on rolling subsample windows for multiple choices of the VAR process lag order p and forecast horizon H parameters. The left-hand (right-hand) side graphs report the results for the forecast horizon $H = 10$ ($H = 30$). Each graph displays the rolling-subsample results for the Total Spillover Index corresponding to the lag order values ranging from $p = 1$ to $p = 10$ (shaded area). The highlighted line corresponds to the lag order $p = 5$. The length of one subsample is 252 days, thus, the subsample start date is approximately one year earlier than the subsample end date (displayed on the x-axis). We plot all three classes of variables entering the underlying VAR processes: log-returns (top row), their volatility (middle row), and their log-volatility (bottom row).

Bibliography

Abad, David, and José Yagüe. 2012. "From PIN to VPIN: An introduction to order flow toxicity." *The Spanish Review of Financial Economics* 10 (2): 74–83.

Acharya, Viral, Robert Engle, and Matthew Richardson. 2012. "Capital shortfall: A new approach to ranking and regulating systemic risks." *American Economic Review* 102 (3): 59–64.

Adrian, Tobias, and Markus K Brunnermeier. 2011. "CoVaR." Technical Report, National Bureau of Economic Research.

Afonso, António, Michael G Arghyrou, George Bagdatoglou, and Alexandros Kontonikas. 2015. "On the time-varying relationship between EMU sovereign spreads and their determinants." *Economic Modelling* 44:363–371.

Agudelo, Diego A, Santiago Giraldo, and Edwin Villarraga. 2015. "Does PIN measure information? Informed trading effects on returns and liquidity in six emerging markets." *International Review of Economics & Finance* 39:149–161.

Ahmad, Wasim, Anil V Mishra, and Kevin J Daly. 2018. "Financial connectedness of BRICS and global sovereign bond markets." *Emerging Markets Review* 37:1–16.

Aktas, Nihat, Eric De Bodt, Fany Declerck, and Herve Van Oppens. 2007. "The PIN anomaly around M&A announcements." *Journal of Financial Markets* 10 (2): 169–191.

Anatolyev, Stanislav, Sergei Seleznev, and Veronika Selezneva. 2018. *Formation of Market Beliefs in the Oil Market*. Economics Institute, Academy of Sciences of the Czech Republic.

Andersen, Torben G, Tim Bollerslev, Francis X Diebold, and Clara Vega. 2003. "Micro effects of macro announcements: Real-time price discovery in foreign exchange." *American economic review* 93 (1): 38–62.

Andersen, Torben G, and Oleg Bondarenko. 2014a. "Reflecting on the VPIN dispute." *Journal of Financial Markets* 17:53–64.

———. 2014b. "VPIN and the flash crash." *Journal of Financial Markets* 17:1–46.

Antonakakis, Nikolaos. 2012. "Exchange return co-movements and volatility spillovers before and after the introduction of euro." *Journal of International Financial Markets, Institutions and Money* 22 (5): 1091–1109.

Antonakakis, Nikolaos, David Gabauer, and Rangan Gupta. 2019. "International monetary policy spillovers: evidence from a time-varying parameter vector autoregression." *International Review of Financial Analysis* 65:101382.

Antonakakis, Nikolaos, and Konstantinos Vergos. 2013. "Sovereign bond yield spillovers in the Euro zone during the financial and debt crisis." *Journal of International Financial Markets, Institutions and Money* 26:258–272.

Arellano, Manuel, et al. 1987. "Computing robust standard errors for within-groups estimators." *Oxford bulletin of Economics and Statistics* 49 (4): 431–434.

Armstrong, Will J, Laura Cardella, and Nasim Sabah. 2017. "Information shocks and liquidity innovations." *Available at SSRN 3003385*.

Bajgrowicz, Pierre, and Olivier Scaillet. 2012. "Technical trading revisited: False discoveries, persistence tests, and transaction costs." *Journal of Financial Economics* 106 (3): 473–491.

Balli, Faruk. 2009. "Spillover effects on government bond yields in euro zone. Does full financial integration exist in European government bond markets?" *Journal of Economics and Finance* 33 (4): 331–363.

Barberis, Nicholas, and Richard Thaler. 2005. *A survey of behavioral finance*. Princeton University Press.

Barndorff-Nielsen, Ole E, Silja Kinnebrock, and Neil Shephard. 2008. "Measuring downside risk-realised semivariance." *CREATES Research Paper*, no. 2008-42.

Baruník, Jozef, and Tobias Kley. 2019. "Quantile coherency: A general measure for dependence between cyclical economic variables." *The Econometrics Journal* 22 (2): 131–152.

Baruník, Jozef, Evžen Kočenda, and Lukáš Vácha. 2016. "Asymmetric connectedness on the US stock market: Bad and good volatility spillovers." *Journal of Financial Markets* 27:55–78.

———. 2017. "Asymmetric volatility connectedness on the forex market." *Journal of International Money and Finance* 77:39–56.

Bash, Ahmad, and Khaled Alsaifi. 2019. "Fear from uncertainty: An event study of Khashoggi and stock market returns." *Journal of Behavioral and Experimental Finance* 23:54–58.

Bessembinder, Hendrik, and Kalok Chan. 1995. "The profitability of technical trading rules in the Asian stock markets." *Pacific-basin finance journal* 3 (2-3): 257–284.

———. 1998. "Market efficiency and the returns to technical analysis." *Financial management*, pp. 5–17.

Bjursell, Johan, James E Gentle, and George HK Wang. 2015. "Inventory announcements, jump dynamics, volatility and trading volume in US energy futures markets." *Energy Economics* 48:336–349.

Bjursell, Johan, George HK Wang, and Hui Zheng. 2017. "VPIN, jump dynamics and inventory announcements in energy futures markets." *Journal of Futures Markets* 37 (6): 542–577.

Brock, William, Josef Lakonishok, and Blake LeBaron. 1992. "Simple technical trading rules and the stochastic properties of stock returns." *The Journal of finance* 47 (5): 1731–1764.

Bu, Hui. 2014. "Effect of inventory announcements on crude oil price volatility." *Energy Economics* 46:485–494.

Campbell, John Y, Andrew W Lo, and A Craig MacKinlay. 2012. *The econometrics of financial markets*. princeton University press.

Chang, Charles, Hazem Daouk, and Albert Wang. 2009. "Do investors learn about analyst accuracy? A study of the oil futures market." *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 29 (5): 414–429.

Chang, Charles, and Emily Lin. 2015. "Cash-futures basis and the impact of market maturity, informed trading, and expiration effects." *International Review of Economics & Finance* 35:197–213.

Chang, Sanders S, Lenisa V Chang, and F Albert Wang. 2014. "A dynamic intraday measure of the probability of informed trading and firm-specific return variation." *Journal of Empirical Finance* 29:80–94.

Chen, Te-Feng, and Huimin Chung. 2007. "The joint estimation of probability of informed trading: The informational role of stock and option volume." In *National Chiao Tung University Working Paper*. Citeseer.

Chen, Yifan, and Huainan Zhao. 2012. "Informed trading, information uncertainty, and price momentum." *Journal of Banking & Finance* 36 (7): 2095–2109.

Cheung, William M, Robin K Chou, and Adrian CH Lei. 2015. "Exchange-traded barrier option and VPIN: Evidence from Hong Kong." *Journal of Futures Markets* 35 (6): 561–581.

Cheung, Yin-Wong, and Menzie David Chinn. 2001. "Currency traders and exchange rate dynamics: a survey of the US market." *Journal of international Money and Finance* 20 (4): 439–471.

Chong, Terence Tai-Leung, Sam Ho-Sum Cheng, and Elfreda Nga-Yee Wong. 2010. "A comparison of stock market efficiency of the BRIC countries." *Technology and Investment* 1 (4): 235.

Christiansen, Charlotte. 2007. "Volatility-spillover effects in European bond markets." *European Financial Management* 13 (5): 923–948.

Claeys, Peter, and Bořek Vašíček. 2014. "Measuring bilateral spillover and testing contagion on sovereign bond markets in Europe." *Journal of Banking & Finance* 46:151–165.

Clare, Andrew, and Ilias Lekkos. 2000. "An analysis of the relationship between international bond markets."

Costantini, Mauro, Matteo Fragetta, and Giovanni Melina. 2014. "Determinants of sovereign bond yield spreads in the EMU: An optimal currency area perspective." *European Economic Review* 70:337–349.

De Bondt, Werner FM, Yaz Gulnur Muradoglu, Hersh Shefrin, and Sotiris K Staikouras. 2008. "Behavioral finance: Quo vadis?" *Journal of Applied Finance (Formerly Financial Practice and Education)* 18, no. 2.

De Bondt, Werner FM, and Richard Thaler. 1985. "Does the stock market overreact?" *The Journal of finance* 40 (3): 793–805.

Diebold, Francis X, and Kamil Yilmaz. 2009. "Measuring financial asset return and volatility spillovers, with application to global equity markets." *The Economic Journal* 119 (534): 158–171.

———. 2012. "Better to give than to receive: Predictive directional measurement of volatility spillovers." *International Journal of forecasting* 28 (1): 57–66.

Diebold, Francis X, and Kamil Yılmaz. 2014. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of econometrics* 182 (1): 119–134.

Diebold, Francis X, and Kamil Yilmaz. 2015. "Trans-Atlantic equity volatility connectedness: US and European financial institutions, 2004–2014." *Journal of Financial Econometrics* 14 (1): 81–127.

Diebold, Francis X, and Kamil Yılmaz. 2015. *Financial and macroeconomic connectedness: A network approach to measurement and monitoring.* Oxford University Press, USA.

Dimpfl, Thomas. 2011. "The impact of US news on the German stock market - An event study analysis." *The Quarterly Review of Economics and Finance* 51 (4): 389–398.

Do, Hung Xuan, Robert Brooks, and Sirimon Treepongkaruna. 2015. "Realized spill-over effects between stock and foreign exchange market: Evidence from regional analysis." *Global Finance Journal* 28:24–37.

Easley, David, Marcos M López de Prado, and Maureen O'Hara. 2010. "Measuring flow toxicity in a high frequency world." *Unpublished Working paper, Cornell University and Tudor Investment Corp.*

———. 2011. "The microstructure of the flash crash: flow toxicity, liquidity crashes, and the probability of informed trading." *The Journal of Portfolio Management* 37 (2): 118–128.

———. 2012. "Flow toxicity and liquidity in a high-frequency world." *The Review of Financial Studies* 25 (5): 1457–1493.

———. 2014. "VPIN and the flash crash: A rejoinder." *Journal of Financial Markets* 17:47–52.

Easley, David, Robert F Engle, Maureen O'Hara, and Liuren Wu. 2008. "Time-varying arrival rates of informed and uninformed trades." *Journal of Financial Econometrics* 6 (2): 171–207.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara. 2002. "Is information risk a determinant of asset returns?" *The journal of finance* 57 (5): 2185–2221.

Easley, David, Nicholas M Kiefer, Maureen O'Hara, and Joseph B Paperman. 1996. "Liquidity, information, and infrequently traded stocks." *The Journal of Finance* 51 (4): 1405–1436.

Easley, David, and Maureen O'Hara. 1987. "Price, trade size, and information in securities markets." *Journal of Financial economics* 19 (1): 69–90.

Ejsing, Jacob, and Jukka Sihvonen. 2009. "Liquidity premia in German government bonds."

Engle, Robert, and Bryan Kelly. 2012. "Dynamic equicorrelation." *Journal of Business & Economic Statistics* 30 (2): 212–228.

Engle, Robert, and Breno Neri. 2010. "The impact of hedging costs on the bid and ask spread in the options market." *Unpublished Working Paper, New York University.*

Fama. 1970. "Efficient capital markets: A review of theory and empirical works." *Journal of Finance* 25 (2): 383–417.

Fama, Eugene F, and Marshall E Blume. 1966. "Filter rules and stock-market trading." *The Journal of Business* 39 (1): 226–241.

Faust, Jon, and Charles H Whiteman. 1997. "General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model to the DGP: A translation and critique." *Carnegie-Rochester Conference Series on Public Policy*, Volume 47. Elsevier, 121–161.

Favero, Carlo, and Alessandro Missale. 2012. "Sovereign spreads in the eurozone: which prospects for a Eurobond?" *Economic Policy* 27 (70): 231–273.

Fernández-Rodríguez, Fernando, Marta Gómez-Puig, and Simón Sosvilla-Rivero. 2015. "Volatility spillovers in EMU sovereign bond markets." *International Review of Economics & Finance* 39:337–352.

———. 2016. "Using connectedness analysis to assess financial stress transmission in EMU sovereign bond market volatility." *Journal of international Financial Markets, institutions and Money* 43:126–145.

Fifield, Suzanne GM, David M Power, and C Donald Sinclair. 2005. "An analysis of trading strategies in eleven European stock markets." *The European Journal of Finance* 11 (6): 531–548.

Finta, Marinela Adriana, and Sofiane Aboura. 2020. "Risk premium spillovers among stock markets: Evidence from higher-order moments." *Journal of Financial Markets* 49:100533.

Frömmel, Michael, and Kevin Lampaert. 2016. "Does frequency matter for intraday technical trading?" *Finance Research Letters* 18:177–183.

Froot, Kenneth A, David S Scharfstein, and Jeremy C Stein. 1992. "Herd on the street: Informational inefficiencies in a market with short-term speculation." *The Journal of finance* 47 (4): 1461–1484.

Gencay, Ramazan. 1999. "Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules." *Journal of International Economics* 47 (1): 91–107.

Glosten, Lawrence R, and Paul R Milgrom. 1985. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *Journal of financial economics* 14 (1): 71–100.

Gómez-Puig, Marta, Simón Sosvilla-Rivero, and María del Carmen Ramos-Herrera. 2014. "An update on EMU sovereign yield spread drivers in times of crisis: A panel data analysis." *The North American Journal of Economics and Finance* 30:133–153.

Gradojevic, Nikola, and Ramazan Gençay. 2013. "Fuzzy logic, trading uncertainty and technical trading." *Journal of Banking & Finance* 37 (2): 578–586.

Greenwood-Nimmo, Matthew, Viet Hoang Nguyen, and Barry Rafferty. 2016. "Risk and return spillovers among the G10 currencies." *Journal of Financial Markets* 31:43–62.

Grimes, Adam. 2012. *The art and science of technical analysis: market structure, price action, and trading strategies*. Volume 544. John Wiley & Sons.

Grossman, Sanford J, and Joseph E Stiglitz. 1980. "On the impossibility of informationally efficient markets." *The American economic review* 70 (3): 393–408.

Hamill, Philip A, Youwei Li, Athanasios A Pantelous, Samuel A Vigne, and James Waterworth. 2021. "Was a deterioration in "connectedness" a leading indicator of the European sovereign debt crisis?" *Journal of International Financial Markets, Institutions and Money*, p. 101300.

Han, Heejoon, Oliver Linton, Tatsushi Oka, and Yoon-Jae Whang. 2016. "The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series." *Journal of Econometrics* 193 (1): 251–270.

Hansen, Peter Reinhard. 2005. "A test for superior predictive ability." *Journal of Business & Economic Statistics* 23 (4): 365–380.

Hendry, David F, et al. 1995. *Dynamic econometrics*. Oxford University Press on Demand.

Heyman, Dries, Koen Inghelbrecht, and Stefaan Pauwels. 2012. "Technical trading rules in emerging stock markets." *Available at SSRN 1998036*.

Hsu, Po-Hsuan, Yu-Chin Hsu, and Chung-Ming Kuan. 2010. "Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias." *Journal of Empirical Finance* 17 (3): 471–484.

Hudson, Robert, Michael Dempsey, and Kevin Keasey. 1996. "A note on the weak form efficiency of capital markets: The application of simple technical trading rules to UK stock prices-1935 to 1994." *Journal of banking & finance* 20 (6): 1121–1132.

Ilmanen, Antti. 1995. "Time-varying expected returns in international bond markets." *The Journal of Finance* 50 (2): 481–506.

Jensen, M, and GA Bennington. 1970. "Random Walks and Technical Theories: Some Additional Evidence." *Journal of Finance (May1960)*, pp. 469–82 Jensen May Journal.

Jensen, Michael C. 1978. "Some anomalous evidence regarding market efficiency." *Journal of financial economics* 6 (2/3): 95–101.

Ji, Qiang, Dayong Zhang, and Jiang-bo Geng. 2018. "Information linkage, dynamic spillovers in prices and volatility between the carbon and energy markets." *Journal of Cleaner Production* 198:972–978.

Jiang, Junhua, Vanja Piljak, Aviral Kumar Tiwari, and Janne Äijö. 2020. "Frequency volatility connectedness across different industries in China." *Finance research letters* 37:101–376.

Kahneman, Daniel, and Amos Tversky. 2013. "Prospect theory: An analysis of decision under risk." In *Handbook of the fundamentals of financial decision making: Part I*, 99–127. World Scientific.

Katusiime, Lorna, Abul Shamsuddin, and Frank W Agbola. 2015. "Foreign exchange market efficiency and profitability of trading rules: Evidence from a developing country." *International Review of Economics & Finance* 35:315–332.

Kaufman, Perry J. 2013. *Trading Systems and Methods,+ Website*. Volume 591. John Wiley & Sons.

Kim, Jae H, Abul Shamsuddin, and Kian-Ping Lim. 2011. "Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data." *Journal of Empirical Finance* 18 (5): 868–879.

Ko, Kuan-Cheng, Shinn-Juh Lin, Hsiang-Ju Su, and Hsing-Hua Chang. 2014. "Value investing and technical analysis in Taiwan stock market." *Pacific-Basin Finance Journal* 26:14–36.

Koop, Gary, M Hashem Pesaran, and Simon M Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of econometrics* 74 (1): 119–147.

Laopodis, Nikiforos T. 2004. "Monetary policy implications of comovements among long-term interest rates." *Journal of International Financial Markets, Institutions and Money* 14 (2): 135–164.

Lee, Charles MC, and Mark J Ready. 1991. "Inferring trade direction from intraday data." *The Journal of Finance* 46 (2): 733–746.

Lee, Chun I, Ming-Shiun Pan, and Y Angela Liu. 2001. "On market efficiency of Asian foreign exchange rates: evidence from a joint variance ratio test and technical trading rules." *Journal of International Financial Markets, Institutions and Money* 11 (2): 199–214.

Lee, Yen-Hsien, Wen-Chien Liu, Chia-Lin Hsieh, et al. 2017. "Informed trading of futures markets during the financial crisis: Evidence from the VPIN." *International Journal of Economics and Finance* 9 (9): 123–132.

Lento, Camillo, Nikola Gradojevic, and CS Wright. 2007. "Investment information content in Bollinger Bands?" *Applied Financial Economics Letters* 3 (4): 263–267.

Leung, Joseph Man-Joe, and Terence Tai-Leung Chong. 2003. "An empirical comparison of moving average envelopes and Bollinger Bands." *Applied Economics Letters* 10 (6): 339–341.

Levich, Richard M, and Lee R Thomas III. 1993. "The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach." *Journal of international Money and Finance* 12 (5): 451–474.

Lim, Kian-Ping, and Robert Brooks. 2011. "The evolution of stock market efficiency over time: A survey of the empirical literature." *Journal of Economic Surveys* 25 (1): 69–108.

Linn, Scott C, and Zhen Zhu. 2004. "Natural gas prices and the gas storage report: Public news and volatility in energy futures markets." *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 24 (3): 283–313.

Lo, Andrew W. 2005. "Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis." *Journal of investment consulting* 7 (2): 21–44.

Maddala, GS, and Hongyi Li. 1996. "15 Bootstrap based tests in financial models." *Handbook of statistics* 14:463–488.

Marshall, Ben R, and Rochester H Cahan. 2005. "Is technical analysis profitable on a stock market which has characteristics that suggest it may be inefficient?" *Research in International Business and Finance* 19 (3): 384–398.

Marshall, Ben R, Rochester H Cahan, and Jared M Cahan. 2008. "Does intraday technical analysis in the US equity market have value?" *Journal of Empirical Finance* 15 (2): 199–210.

———. 2010. "Technical analysis around the world." *Available at SSRN 1181367*.

McCauley, Robert N, et al. 1999. "The euro and the liquidity of European fixed income markets." *Market Liquidity: Research Findings and Selected Policy Implications, BIS CGFS Publication*, no. 11.

Menkveld, Albert J, and Bart Zhou Yueshen. 2013. "Anatomy of the flash crash." *Available at SSRN 2243520*.

Metghalchi, Massoud, Juri Marcucci, and Yung-Ho Chang. 2012. "Are moving average trading rules profitable? Evidence from the European stock markets." *Applied Economics* 44 (12): 1539–1559.

Murphy, John J. 1999. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

Neely, Christopher J, and Paul A Weller. 2003. "Intraday technical trading in the foreign exchange market." *Journal of International Money and Finance* 22 (2): 223–237.

Neely, Christopher J, Paul A Weller, and Joshua M Ulrich. 2009. "The adaptive markets hypothesis: evidence from the foreign exchange market." *Journal of Financial and Quantitative Analysis* 44 (2): 467–488.

Ntim, Collins G, Kwaku K Opong, Jo Danbolt, and Frank Senyo Dewotor. 2011. "Testing the weak-form efficiency in African stock markets." *Managerial Finance*.

Osler, Carol L. 2000. "Support for resistance: technical analysis and intraday exchange rates." *Economic Policy Review* 6, no. 2.

Park, Cheol-Ho, and Scott H Irwin. 2004. "The profitability of technical analysis: A review."

———. 2007. "What do we know about the profitability of technical analysis?" *Journal of Economic surveys* 21 (4): 786–826.

Park, Jung Soo, Chris Heaton, et al. 2014. "Technical trading rules in Australian financial markets."

Pelizzon, Loriana, Marti G Subrahmanyam, Davide Tomio, and Jun Uno. 2014. "Limits to arbitrage in sovereign bonds: Price and liquidity discovery in high-frequency quotedriven markets." *The 41th European Finance Association Annual Meeting (EFA 2014)*.

Pesaran, H Hashem, and Yongcheol Shin. 1998. "Generalized impulse response analysis in linear multivariate models." *Economics letters* 58 (1): 17–29.

Politis, Dimitris N, and Joseph P Romano. 1994. "The stationary bootstrap." *Journal of the American Statistical association* 89 (428): 1303–1313.

Pring, Martin J. 2021. Technical Analysis Explained: The Successful Investor's to Spotting investment trends turning points.

Schulmeister, Stephan. 2009. "Profitability of technical stock trading: Has it moved from daily to intraday data?" *Review of Financial Economics* 18 (4): 190–201.

Sewell, Martin. 2011. "History of the efficient market hypothesis." *Rn* 11 (04): 04.

Sibbertsen, Philipp, Christoph Wegener, and Tobias Basse. 2014. "Testing for a break in the persistence in yield spreads of EMU government bonds." *Journal of Banking & Finance* 41:109–118.

Sullivan, Ryan, Allan Timmermann, and Halbert White. 1999. "Data-snooping, technical trading rule performance, and the bootstrap." *The journal of Finance* 54 (5): 1647–1691.

Tabak, Benjamin M, and Eduardo JA Lima. 2009. "Market efficiency of Brazilian exchange rate: Evidence from variance ratio statistics and technical trading rules." *European Journal of Operational Research* 194 (3): 814–820.

Tan, Xueping, Kavita Sirichand, Andrew Vivian, and Xinyu Wang. 2020. "How connected is the carbon market to energy and financial markets? A systematic analysis of spillovers and dynamics." *Energy Economics* 90:104870.

Urquhart, Andrew, and Frank McGroarty. 2014. "Calendar effects, market conditions and the Adaptive Market Hypothesis: Evidence from long-run US data." *International Review of Financial Analysis* 35:154–166.

Vega, Clara. 2006. "Stock price reaction to public and private information." *Journal of Financial Economics* 82 (1): 103–133.

Wei, Wang Chun, Dionigi Gerace, and Alex Frino. 2013. "Informed trading, flow toxicity and the impact on intraday trading factors." *Australasian Accounting, Business and Finance Journal* 7 (2): 3–24.

Weng, Pei-Shih, Ming-Hung Wu, Miao-Ling Chen, and Wei-Che Tsai. 2017. "An empirical analysis of the dynamic probability of informed institutional trading: evidence from the Taiwan futures exchange." *Journal of Futures Markets* 37 (9): 865–891.

White, Halbert. 2000. "A reality check for data snooping." *Econometrica* 68 (5): 1097–1126.

Wu, Kesheng, E Bethel, Ming Gu, David Leinweber, and Oliver Ruebel. 2013. "A big data approach to analyzing market volatility." *Algorithmic Finance* 2 (3-4): 241–267.

Yamamoto, Ryuichi. 2012. "Intraday technical analysis of individual stocks on the Tokyo Stock Exchange." *Journal of Banking & Finance* 36 (11): 3033–3047.

Ye, Shiyu, and Berna Karali. 2016. "The informational content of inventory announcements: Intraday evidence from crude oil futures market." *Energy Economics* 59:349–364.

Yi, Shuyue, Zishuang Xu, and Gang-Jin Wang. 2018. "Volatility connectedness in the cryptocurrency market: Is Bitcoin a dominant cryptocurrency?" *International Review of Financial Analysis* 60:98–114.

Yildiz, Serhat, Bonnie Van Ness, and Robert Van Ness. 2020. "VPIN, liquidity, and return volatility in the US equity markets." *Global Finance Journal* 45:100479.

Yu, Hao, Gilbert V Nartea, Christopher Gan, and Lee J Yao. 2013. "Predictive ability and profitability of simple technical trading rules: Recent evidence from Southeast Asian stock markets." *International Review of Economics & Finance* 25:356–371.

Zhang, Dayong. 2017. "Oil shocks and stock markets revisited: Measuring connectedness from a global perspective." *Energy Economics* 62:323–333.

Zhou, Jian, and Jin Man Lee. 2013. "Adaptive market hypothesis: evidence from the REIT market." *Applied Financial Economics* 23 (21): 1649–1662.